

# SPAM FILTER

- A **program** used to **detect** unsolicited, **unwanted** and **virus-infected emails** and prevent those messages from getting to a user's inbox.
- Simplest and earliest versions of spam filtering, like the one used by **Microsoft's Hotmail**, was set to watch out for particular words in the subject lines of messages.
- An email was **excluded from the user's inbox** whenever the filter **recognized** one of the **specified words**.
- It is not effective and often omits perfectly legitimate messages, called **false positives**, while letting actual spam messages through.

# SPAM FILTER

- More sophisticated programs, such as Bayesian filters and other heuristic filters, identify spam messages by recognizing suspicious word patterns or word frequency.
- There are many different types of spam filters.
- Blocklist filters: it block spam emails from senders that have been added to a **comprehensive spammers list**. It is updated frequently to keep up with spammers who change their email addresses relatively quickly.
- Content filters: it examine the contents of each email and use that information to decide whether it is spam or not. It tend to work because spam email content is often predictable, offering deals, promoting explicit content or targeting basic human feelings, such as desire and fear.

# SPAM FILTER

- **Header filters:** It analyze email **headers** to determine if they originated from a legitimate source.
- **Language filters:** Spammers often target people worldwide and, sometimes, send emails from geographic areas where the language is different from the recipient's native language.
- **Rule-based filters:** It enable users to establish specific rules and apply them to all incoming emails. Whenever content matches one of the rules, it automatically forwards the email to a spam folder. The rules can be **specific words or phrases** in the message or header.

# RECOMMENDER SYSTEM

- It is **designed to recommend things** to the user based on many different factors.
- It **predict the most likely product** that the **users** are most likely to **purchase** and are of interest to.
- Companies like [Netflix](#), Amazon, etc. use recommender systems to help their **users to identify the correct product** or **movies** for them.
- Deals with a large volume of information present by filtering the most important information based on the data provided by a user and other factors that take care of the user's preference and interest.
- It finds out the **match between user and item** and computes the **similarities between users and items** for recommendation.

# BENEFITS OF RECOMMENDER SYSTEM

- Benefits users in finding items of their interest.
  - Help item providers in delivering their items to the right user.
  - Identity products that are most relevant to users.
  - Personalized content.
  - Help websites to improve user engagement.
- Recommendation for movies, books, news, articles, jobs, advertisements, etc.

# TYPES OF RECOMMENDER SYSTEM

## Popularity-Based Recommendation System

- It works on the **principle of popularity** and give information about anything **which is in trend**.
- **E.g.**, if a product is **often purchased by most people** then the system will get to know that that product is most popular so for every new user who just signed it, the system will recommend that product to that user also and chances becomes high that the new user will also purchase that.
- It does not suffer from **cold start problems** which means on day 1 of the business also it can recommend products on various different filters.

## **Merits**

- There is no need for the user's historical data.
- Not personalized
- The system would recommend the same sort of products/movies which are solely based upon popularity to every other user.

# TYPES OF RECOMMENDER SYSTEM

## Classification Model

- The model that uses **features of both products as well as users** to predict whether a user will like a product or not.
- The output can be either 0 or 1. If the user likes it then 1 and vice-versa.

## Limitations of Classification Model

- It is a rigorous task to collect a high volume of information about different users and also products.
- Also, if the collection is done then also it can be difficult to classify.
- Flexibility issue.

# TYPES OF RECOMMENDER SYSTEM

## Content-Based Recommendation System

- It works on the principle of similar content. If a user is watching a movie, then the system will check about other movies of similar content or the same genre.
- To check the similarity, system computes distances between products. One plus 7 and One plus 7T both have 8Gb ram and 48MP primary camera.
- To check similarity between products, Euclidean distance is calculated. Euclidean distance between (7T,7) is 0 whereas Euclidean distance between (7pro,7) is 4 which means one plus 7 and one plus 7T have similarities in them whereas one plus 7Pro and 7 are not similar products.
- For computing the similarity between numeric data, Euclidean distance is used, for textual data, cosine similarity is calculated and for categorical data, Jaccard similarity is computed.



# TYPES OF RECOMMENDER SYSTEM

- **Cosine Similarity:** Cosine of the angle between the two vectors of the item, vectors of A and B is calculated for computing similarity. If the vectors are closer, then small will be the angle and large will be the cosine

$$\text{Similarity}(X,Y) = \frac{X.Y}{|X| \times |Y|}$$

- **Jaccard Similarity:** Users who have rated item A and B divided by the total number of users who have rated either A or B gives us the similarity. It is used for comparing the similarity.

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|}$$

# TYPES OF RECOMMENDER SYSTEM

## Merits

- There is **no requirement for much of the user's data.**
- A content-based recommender engine **does not depend on the user's data, so even if a new user comes in,** we can recommend the user as long as we have the user data to build his profile.
- **It does not suffer from a cold start.**

## Demerits

- Items data should be in good volume.
- **Features should be available to compute the similarity.**

# TYPES OF RECOMMENDER SYSTEM

## Collaborative Filtering

- It **work on the similarity between different users and also items** that are widely used as an e-commerce website and also online movie websites.
- It checks the taste of similar users and does recommendations.
- The **similarity is not restricted to the taste of the user** moreover there can be consideration of **similarity between different items also**. The system will give more efficient recommendations if we have a large volume of information about users and items.
- It is found that both David and Peter have similar tastes so David interest is recommended to Peter and vice versa.
- Mainly, there are two approaches used in collaborative filtering.

# TYPES OF RECOMMENDER SYSTEM

## a.) User-based nearest-neighbor collaborative filtering

- The system **finds out the users who have the same sort of taste of purchasing products** and similarity between users is computed based upon the purchase behavior. User A and User C are similar because they have purchased similar products.

## b.) Item-based nearest-neighbor collaborative filtering

- The system **checks the items that are similar to the items the user bought**. The similarity between different items is computed based on the items and not the users for the prediction. Users X and Y both purchased items A and B so they are found to have similar tastes.

# TYPES OF RECOMMENDER SYSTEM

## Limitations

- **Enough users required to find a match.**
- Even if there are many users and many items that are to be recommended often, problems can arise of user and rating matrix to be sparse and will become challenging to find out about the users who have rated the same item.
- The problem in recommending items to the user due to sparsity problems.