



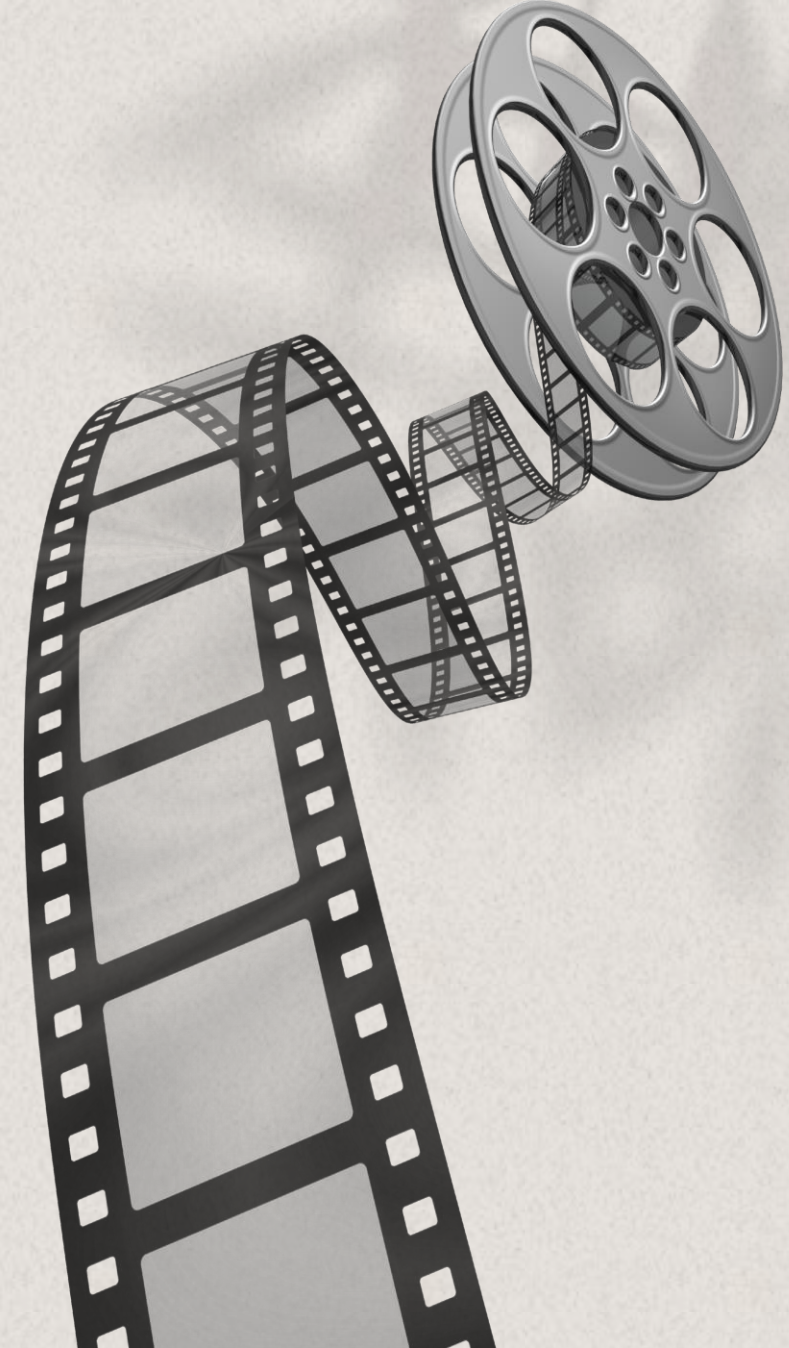
# **Collaborative Filtering Based Recommender & Hybrid Recommender System**

# **Team**

**Swathi Gurijala(C0790294)**

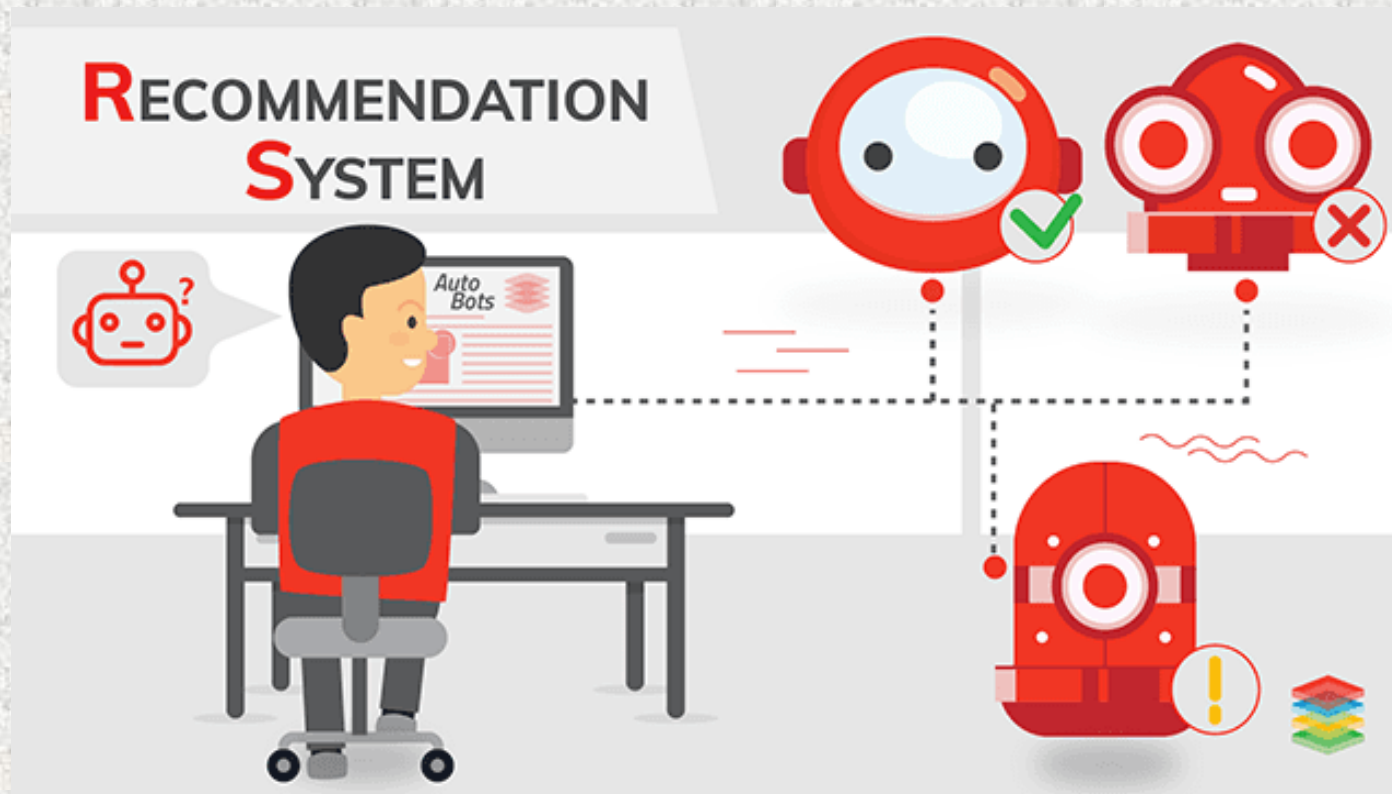
**Varadharajan Kalyanaraman(C0793756)**

**Vignesh Kumar Murugananthan(C0793760)**



# Recommender Systems

---



- Recommender systems are algorithms aimed at suggesting relevant items to users (items being movies to watch, text to read, products to buy or anything else depending on industries) services.



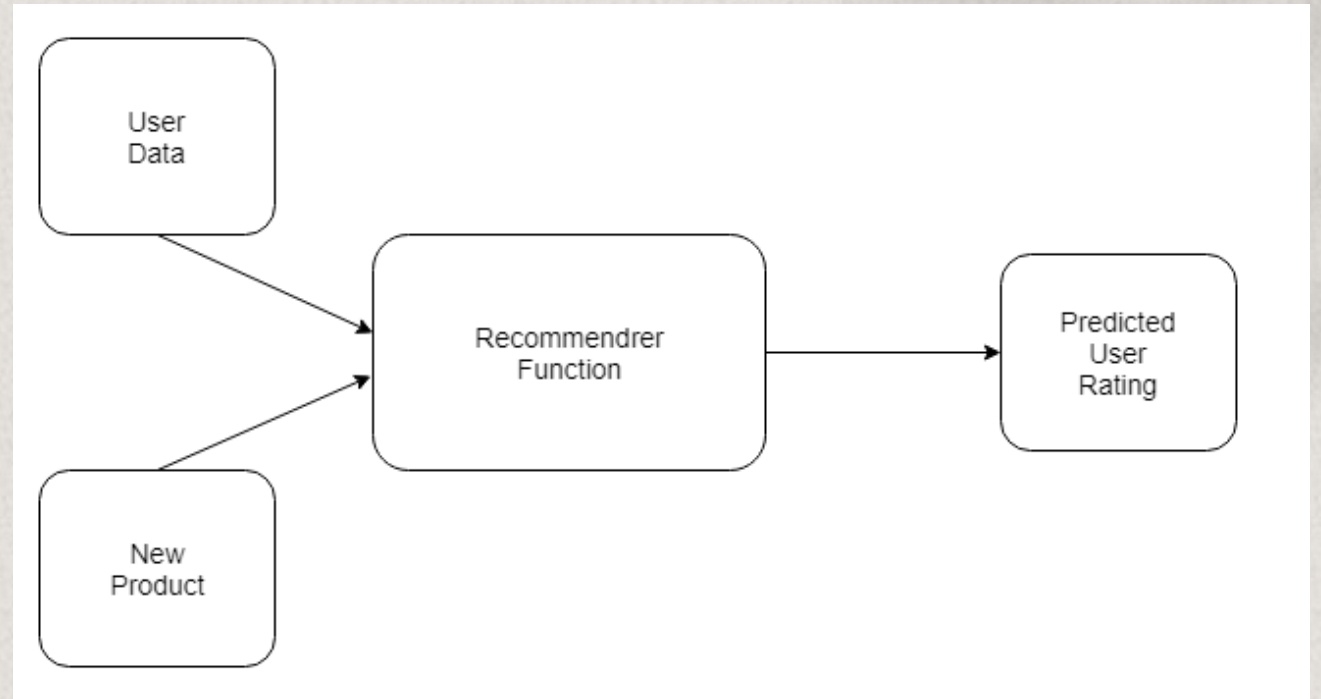
# Overview

## UNDERSTANDING RELATIONSHIPS

- User-Product Relationship
- Product-Product Relationship
- User-User Relationship

## DATA & RECOMMENDER SYSTEMS

- User Behavior Data
- User Demographic Data
- Product Attribute Data

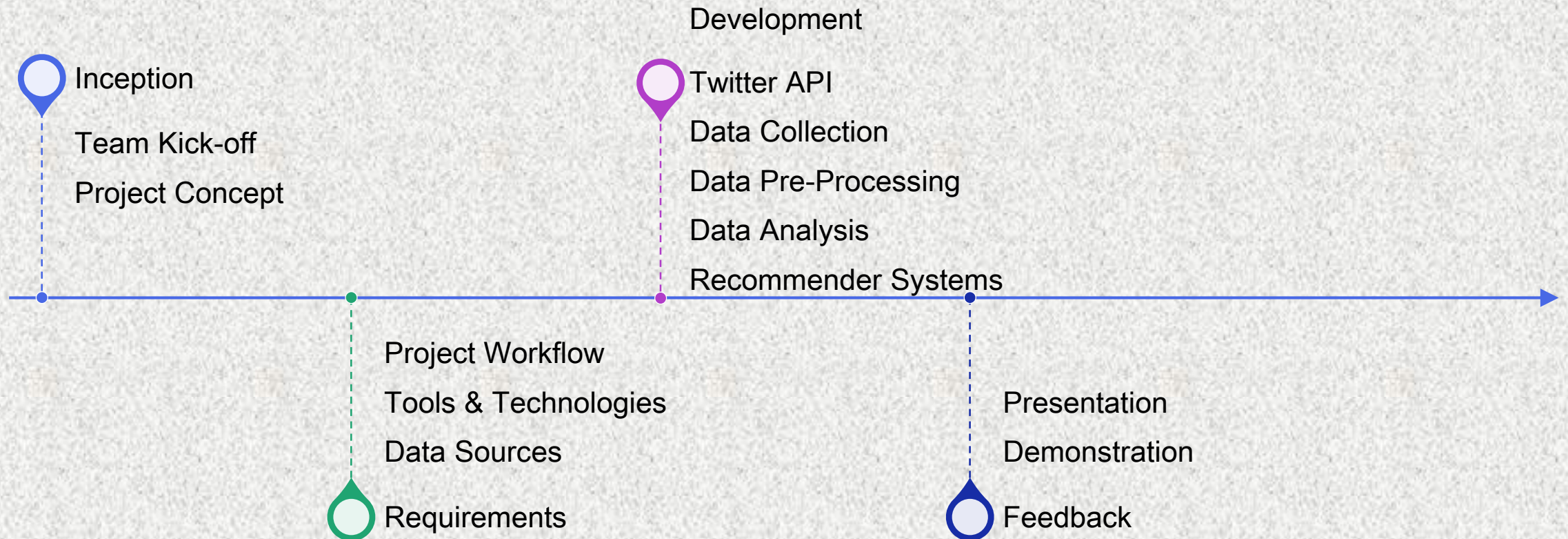


A Simple Layout

# HOW DO WE PROVIDE DATA FOR RECOMMENDER SYSTEMS?

- **Explicit Ratings**
- **Implicit Ratings**
- **Product Similarity (Item-Item Filtering)**
- **User Similarity (User-User Filtering)**

# Project Workflow



# Dataset Description:

---

- The dataset is from the Movie Lens dataset collected by the Group Lens Research Project at the University of Minnesota.

This data set consists of:

- 100,000 ratings (1-5) from 943 users on 1682 movies.
- Each user has rated at least 20 movies.



**u.data** -- The full 'u' data set, 100000 ratings by 943 users on 1682 items. Each user has rated at least 20 movies. Users and items are numbered consecutively from 1. The data is randomly ordered.

**u.item** -- Information about the items (movies)

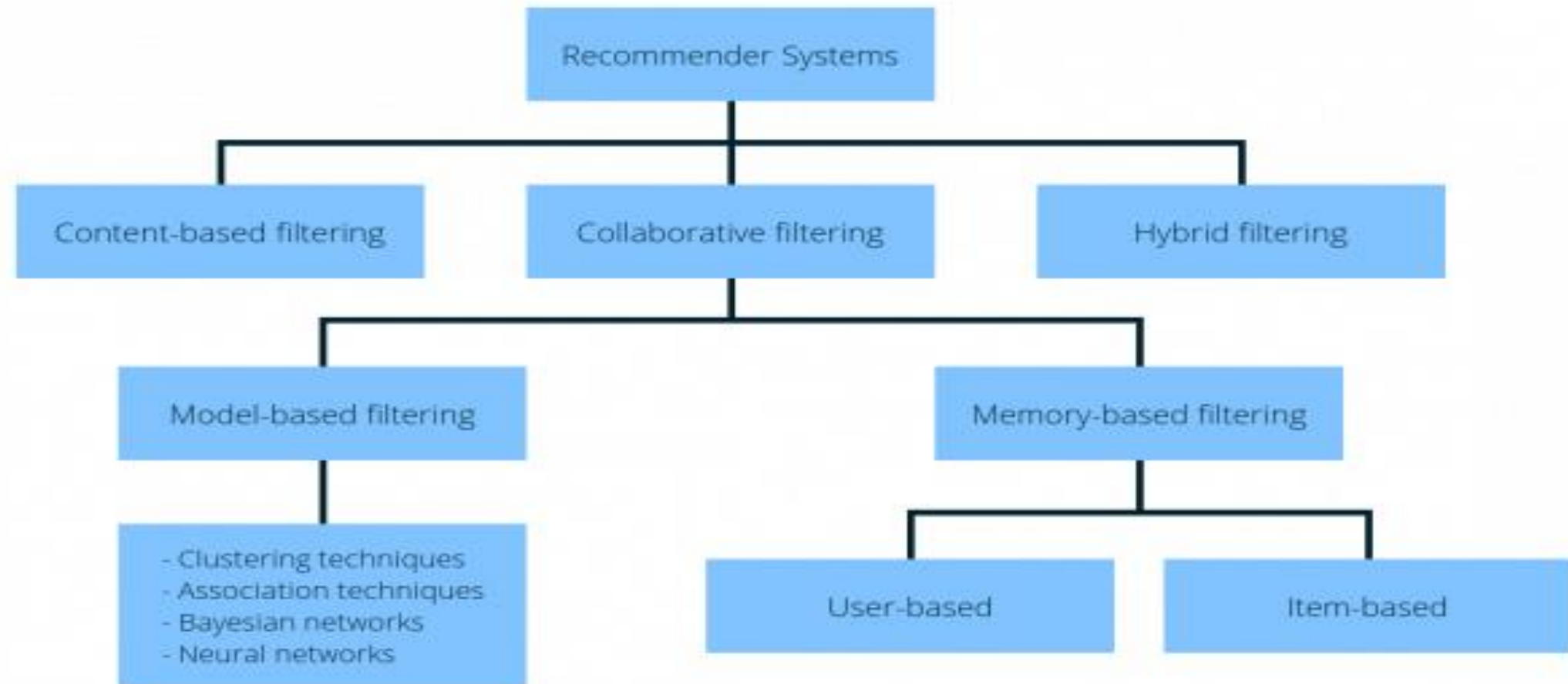
**u.genre** -- A list of the genres.



# Introduction

## Recommender Systems Types

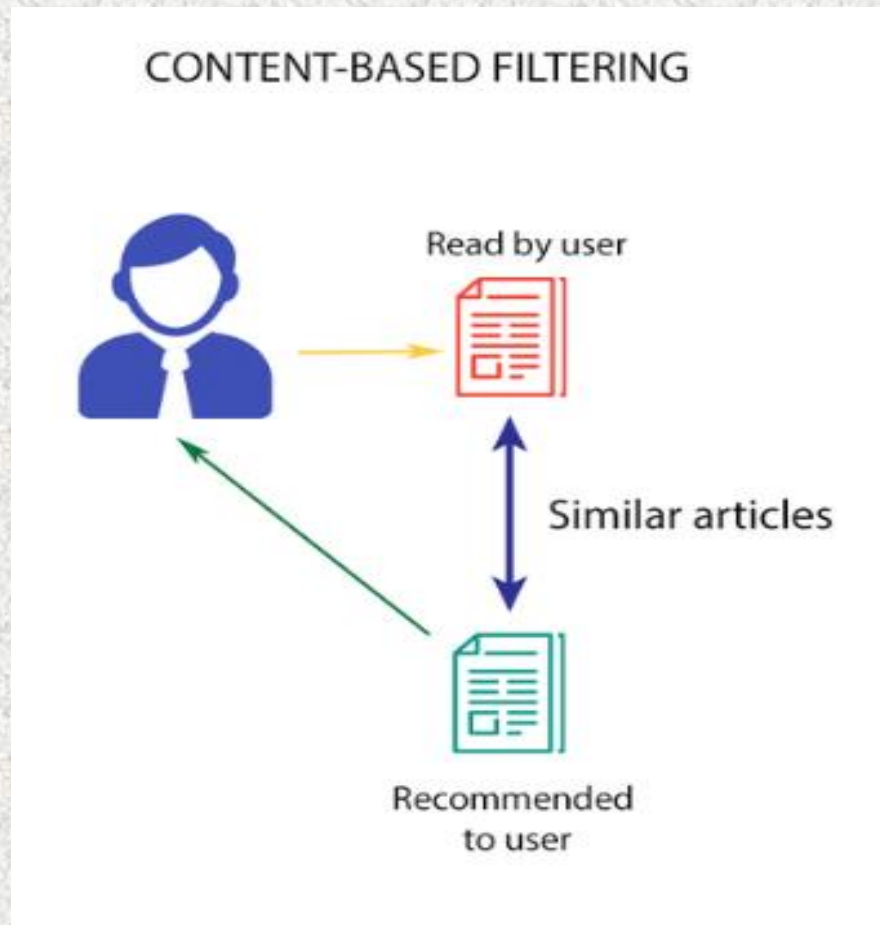
---





# Content Based Filtering

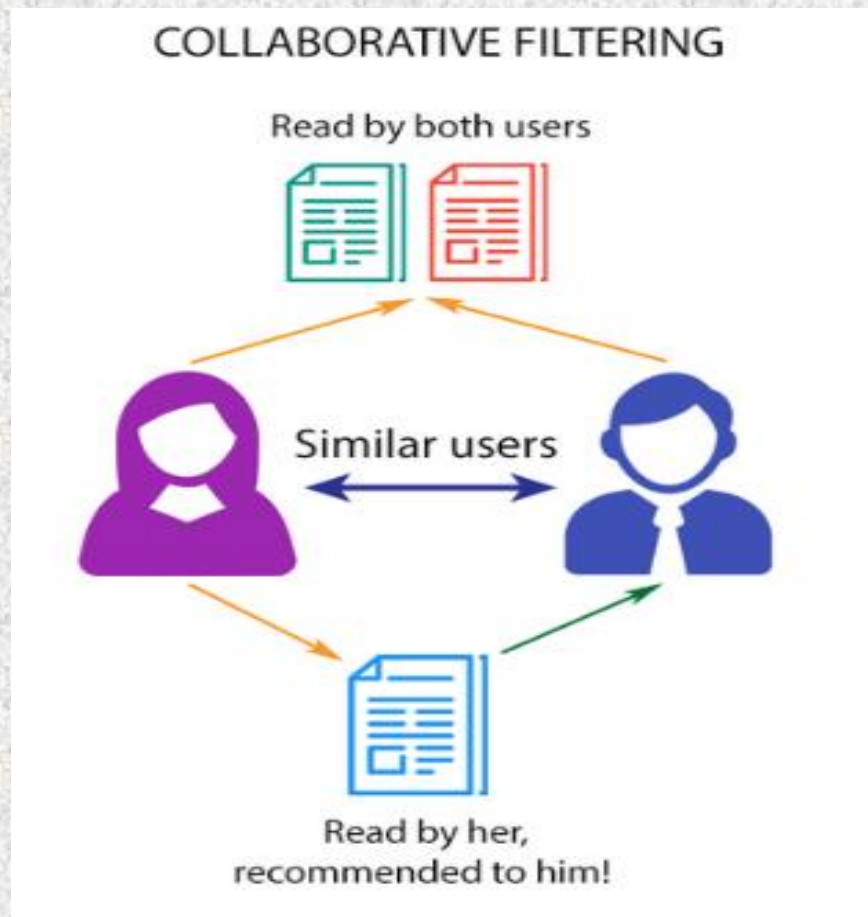
---



- Content-based Filtering is a Machine Learning technique that uses similarities in features to make decisions.
- Algorithms designed to advertise or recommend things to users based on knowledge accumulated about the user.

# Collaborative Based Filtering

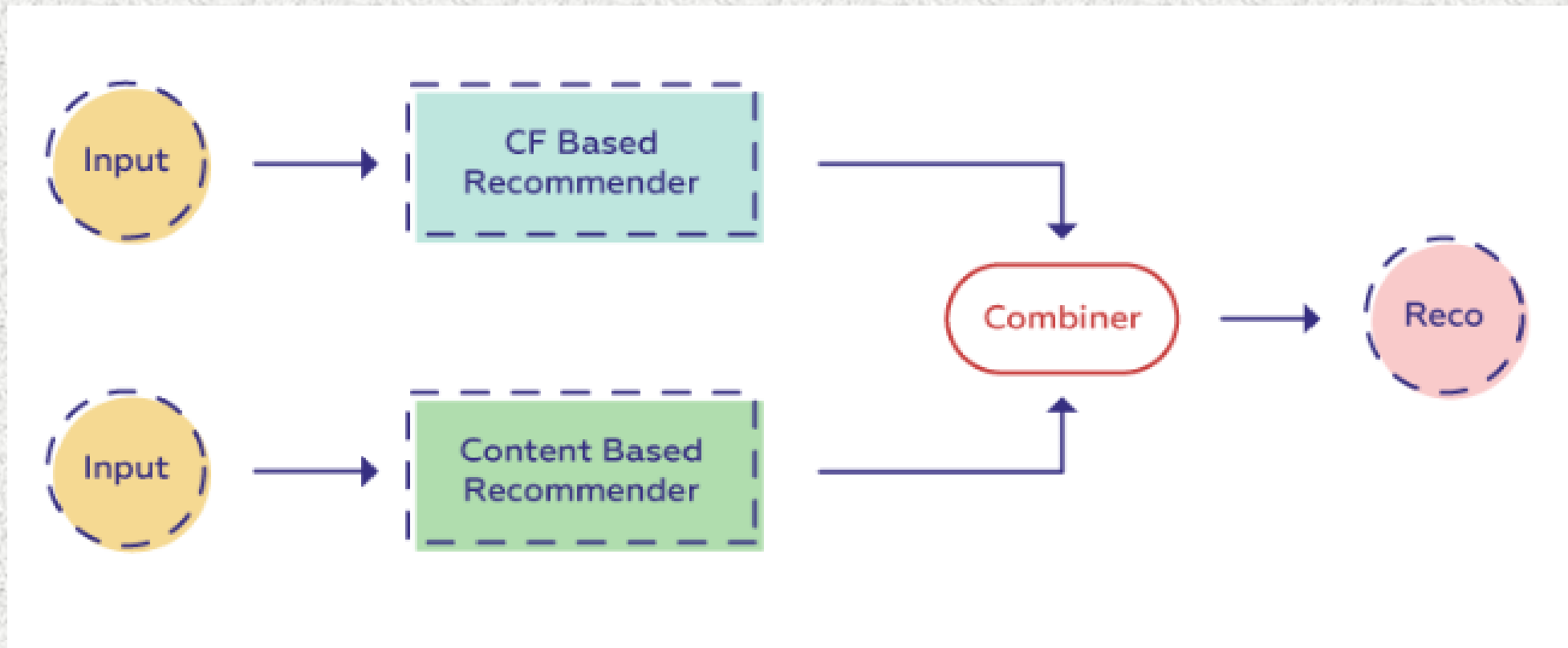
---



- In the newer, narrower sense, collaborative filtering is a method of making automatic predictions (filtering) about the interests of a user by collecting preferences or taste information from many users.
- There are two types –
  - Model Based
  - Memory Based

# Hybrid Filtering

---



# Related Study



- We have Reviewed more than 20 papers that is based on content based and collaborative filtering systems with matrix factorization
- Most used dataset.





# Methodology



# Twitter Data collection and cleaning



- Tweepy package
- Tweet collection on the movie names as hashtags
- Clean the tweets for inconsistency and irregularity
- Make them ready for Sentiment Analysis.



# Matrix Factorization

The diagram shows the equation: Rating Matrix = User Matrix x Item Matrix. Each matrix is a 4x4 grid with rows labeled A, B, C, D and columns labeled W, X, Y, Z.

**Rating Matrix**

	W	X	Y	Z
A		4.5	2.0	
B	4.0		3.5	
C		5.0		2.0
D		3.5	4.0	1.0

**User Matrix**

A	1.2	0.8
B	1.4	0.9
C	1.5	1.0
D	1.2	0.8

**Item Matrix**

W	X	Y	Z
1.5	1.2	1.0	0.8
1.7	0.6	1.1	0.4

- Matrix factorization is a way to generate latent features when multiplying two different kinds of entities.
- Collaborative filtering is the application of matrix factorization to identify the relationship between items' and users' entities.



# Sentiment Score Calculation



Negative



Neutral



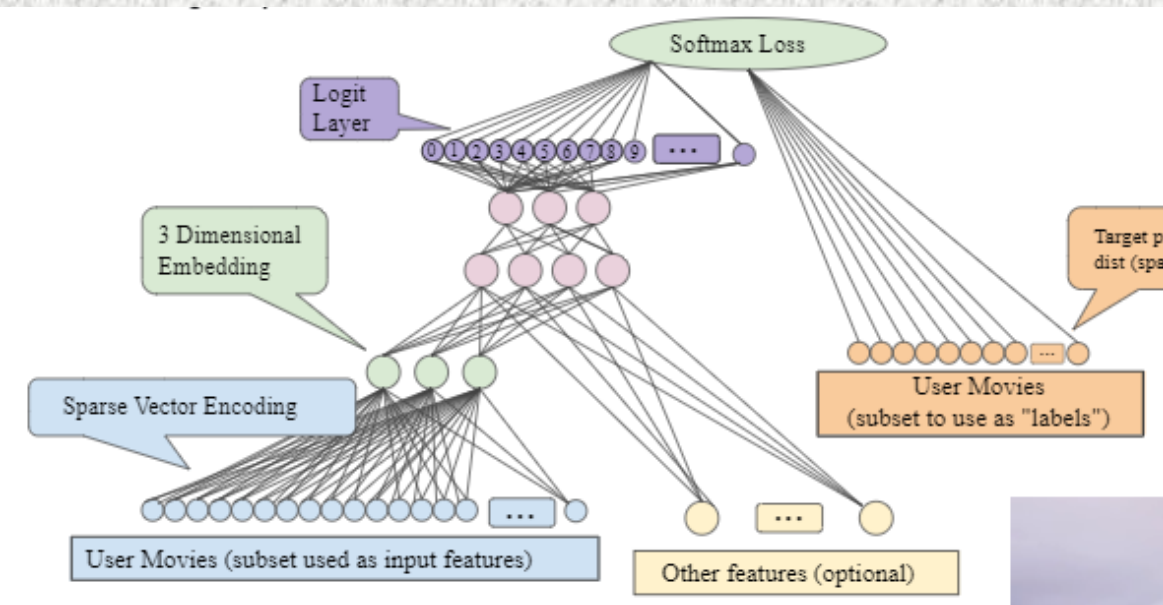
Positive



- For Sentiment and Emotion Analysis, we considered many different tools but ended up selecting VADER.
- VADER (or Valence Aware Dictionary and sEntiment Reasoner) is a binary Sentiment Analysis tool using a dictionary approach, containing 7.518 uni-grams including punctuation, slang words, initialisms, acronyms and emoticons.

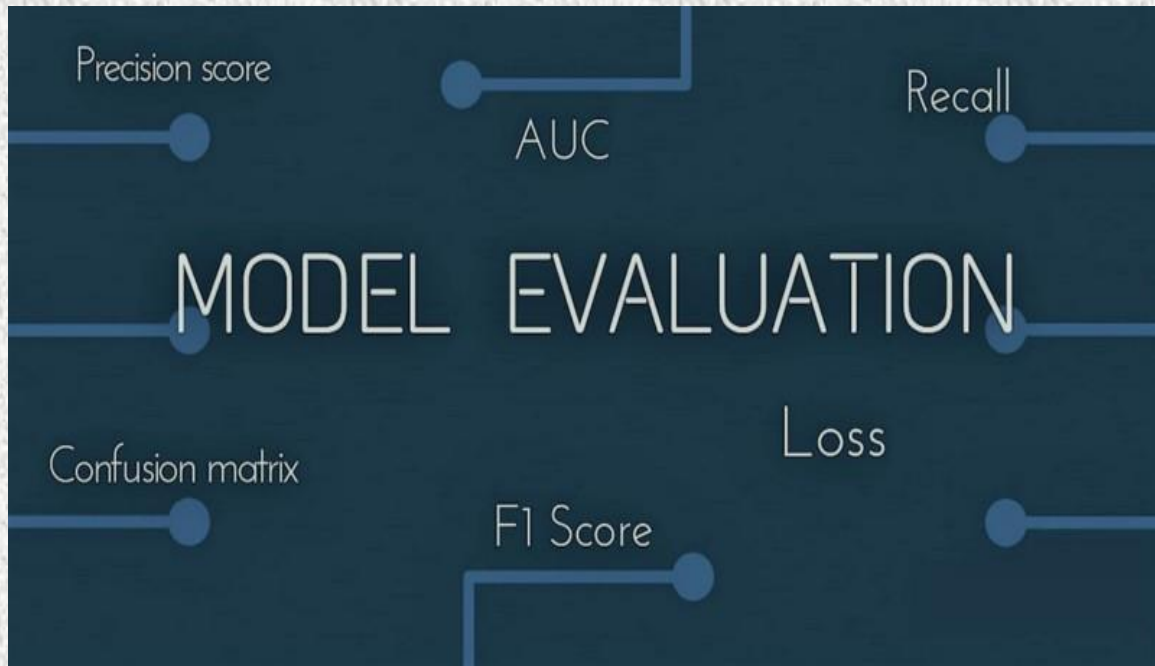


# Neural Embedding Layer



- Neural network embeddings are *learned low-dimensional representations of discrete data as continuous vectors*. These embeddings overcome the limitations of traditional encoding methods and can be used for purposes such as *finding nearest neighbors, input into another model, and visualizations*.
- *Word embeddings are a representation of the \*semantics\* of a word, efficiently encoding semantic information that might be relevant to the task at hand.*

# Evaluation Metrics



- RMSE
- MAE
- PRECISION
- RECALL
- F1-SCORE

# Results and Analysis

## Neural Embedding Metrics

```
Results of sklearn.metrics:  
MAE: 0.7404742048501969  
MSE: 0.8746793064441374  
RMSE: 0.9352429130681169  
R-Squared: 0.3059656687608454  
Precision 0.12087912087912088  
Recall: 0.7534246575342466  
F1-score: 0.20833333333333331
```

- In the newer, narrower sense, collaborative filtering is a method of making automatic predictions (filtering) about the interests of a user by collecting preferences or taste information from many users.

## Matrix Factorization Metrics

```
RMSE: 0.4775128762588308  
MAE: 0.1198866174513692  
Precision: 0.9818181818181818  
Recall: 0.2465753424657534  
F1 Score 0.3941605839416058
```

# Code Walkthrough

```
public FragmentTabHost(Context context) {  
    // Note that we call through to the version that takes a AttributeSet  
    // because the simple Context construct can result in a new object  
    super(context, null);  
    initFragmentTabHost(context, null);  
}
```

```
public FragmentTabHost(Context context, AttributeSet attrs){  
    super(context, attrs);  
}
```



спасибо  
danke 謝謝  
ngiyabonga  
teşekkür ederim  
dank je  
gracias  
tapadh leat  
moichchakkeram  
go raibh maith agat  
arigatō  
takk  
dakujem  
merci  
ευχαριστώ  
terima kasih  
감사합니다  
sagolun  
sukriya  
kop khun krap  
grazie  
brigado  
bedankt  
dziękuję  
hvala  
maunuru