fractal

Team SPY-AI
Gents Presents

# Case Study on Recommendation System of

&

amazon prime

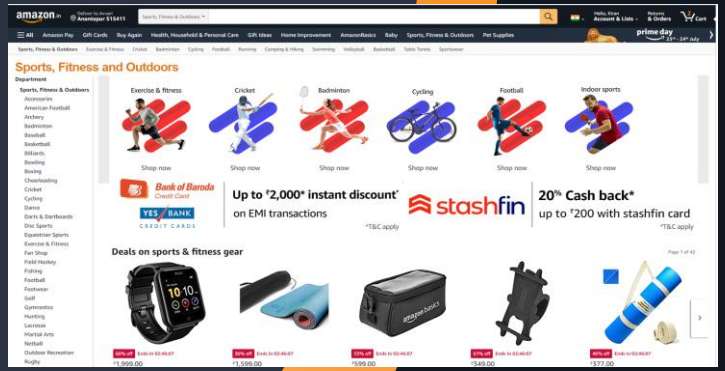
# Business Problem-Introduction

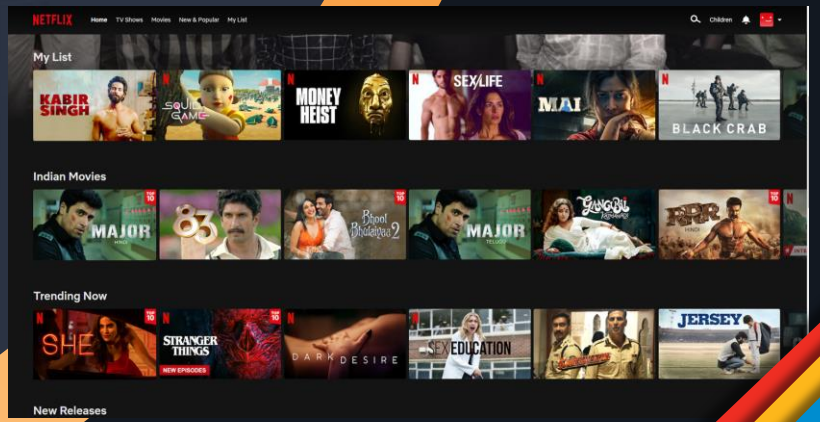
## How To Give The Personalised Experience And Recommendations for a User Based on his/her Interests.

- Benefitting users to find their own interests
- Helping the product providers to deliver to the right user
- Providing items that are most relevant to the user
- Predicting the products that user might use
- Making mood or Time based recommendations for user

# Algorithms Involved

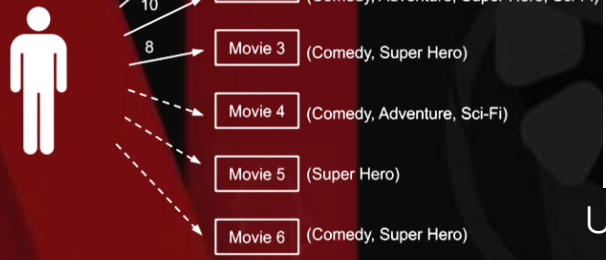
- A9 Algorithm (Specially for amazon)
- Netflix Recommendation Engine(NRE)
  - Content Based Filtering
    - TF-IDF
    - Classifiers(ANN, Naive Bayes)
  - Collaborative Filtering
    - Matrix Factorization
    - Nearest Neighbours
- Multinomial VAE



# Maths Behind The Algorithms:

## Content Based Filtering:




| Movie | Rate |
|---|---|
| 1 | 2 |
| 2 | 10 |
| 3 | 8 |

User Rating Matrix

T

| | Comedy | Adventure | Super Hero | Sci-Fi |
|---|---|---|---|---|
| Movie 1 | 0 | 1 | 1 | 0 |
| Movie 2 | 1 | 1 | 1 | 1 |
| Movie 3 | 1 | 0 | 1 | 0 |

Movie Genre Matrix

| | Comedy | Adventure | Super Hero | Sci-Fi |
|---|---|---|---|---|
| User | 18 | 12 | 20 | 10 |

User Profile

| | Comedy | Adventure | Super Hero | Sci-Fi |
|---|---|---|---|---|
| User | 0.3 | 0.2 | 0.33 | 0.16 |

Normalised User Profile

# Maths Behind The Algorithms:

fractaloo

| | Comedy | Adventure | Super Hero | Sci-Fi |
|---|---|---|---|---|
| User | 0.3 | 0.2 | 0.33 | 0.16 |

Normalised User Profile

×

T

| | Comedy | Adventure | Super Hero | Sci-Fi |
|---|---|---|---|---|
| Movie 4 | 1 | 1 | 0 | 1 |
| Movie 5 | 0 | 0 | 1 | 0 |
| Movie 6 | 1 | 0 | 1 | 0 |

Movie Matrix

- Highest Weight will be Recommended.
- Calculation of rating using streaming time, Times watched, Search history...etc
- Disadvantage of this is the genre which user has never watched will not be in their profile.

| | |
|---|---|
| Movie 4 | 0.66 |
| Movie 5 | 0.33 |
| Movie 6 | 0.63 |

Final Recommendation Matrix

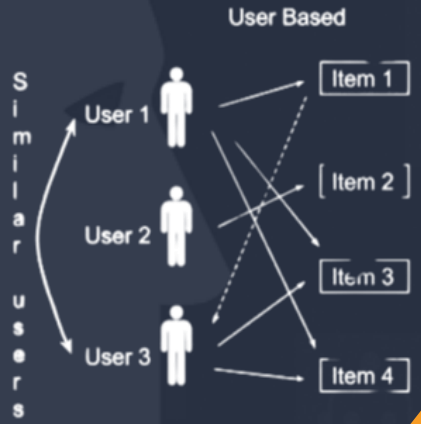
# Maths Behind The Algorithms

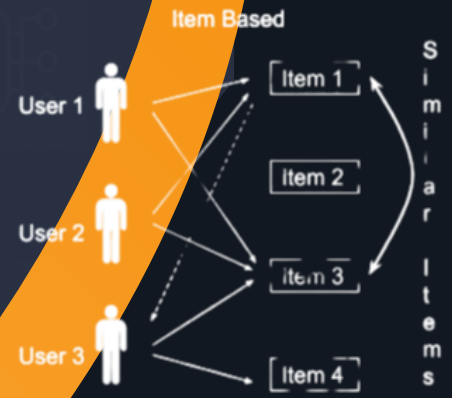
## Collaborative Filtering:

- User based or Item based, identify the similarity/Correlation between them
- By using several distance and similarity measurements like Euclidean distance, Pearson Correlation, Cosine Similarity etc
- Consider the Table:

| | User 1 | User 2 | User 3 |
|---|---|---|---|
| Active User | 0.4 | 0.9 | 0.7 |

- These numbers represent the similarity weights or the proximity of the active user to the other users in the dataset



# Maths Behind The Algorithms:

fractaloo

| | Movie 1 | Movie 2 | Movie 3 | Movie 4 | Movie 5 |
|---|---|---|---|---|---|
| User 1 | 9 | 6 | 8 | 4 | - |
| User 2 | 2 | 10 | 6 | - | 8 |
| User 3 | 5 | 9 | - | 10 | 7 |
| Active User | ?? | 10 | 7 | 8 | ?? |

Users Rating Matrix

| | Movie 1 | Movie 5 |
|---|---|---|
| User 1 | 9 | - |
| User 2 | 2 | 8 |
| User 3 | 5 | 7 |

×

| | Similarity Index |
|---|---|
| User 1 | 0.4 |
| User 2 | 0.9 |
| User 3 | 0.7 |

=

| | Movie 1 | Movie 5 |
|---|---|---|
| User 1 | 3.6 | - |
| User 2 | 1.8 | 7.2 |
| User 3 | 3.5 | 4.9 |

Weighted Rating Matrix

| | Movie 1 | Movie 5 | Weighted Sum |
|---|---|---|---|
| Users 1,2,3 | 3.6+1.8+3.5 = 8.9 | - | 0.9+0.7+0.4 = 2.0 |
| Users 2,3 | - | 7.2+4.9 = 12.1 | 0.9+0.7 = 1.6 |

Weighted Recommendation Matrix

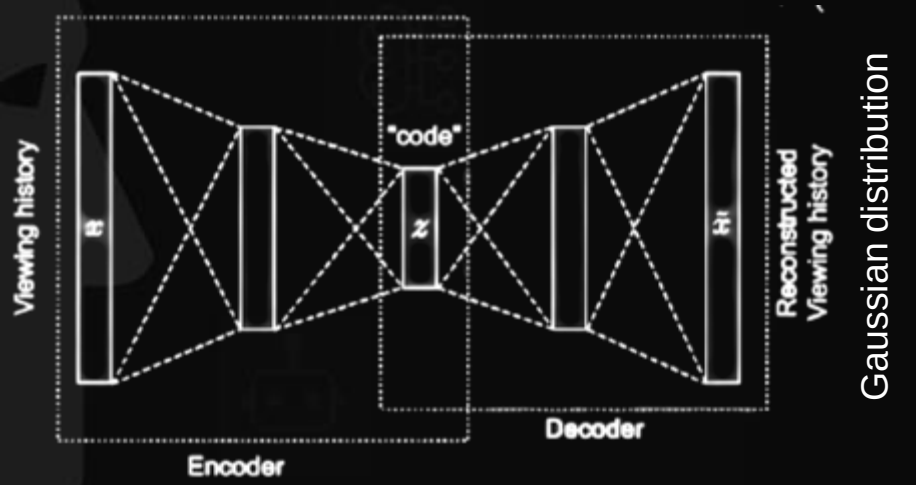
| | Movie 1 | Movie 5 |
|---|---|---|
| Active User | 8.9/2.0 = 4.45 | 12.1/1.6 = 7.5 |

Final Recommendation Matrix

# Factors for Good Customer Experience

## Collaborative Ranking

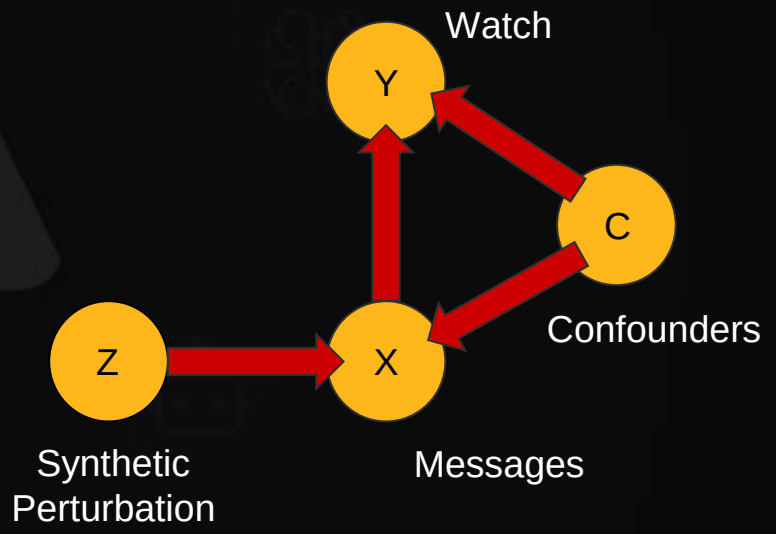
- Ranking movies based on prediction of what user wants to watch
- Aspirational star rating
- User-product matrix
- Multinomial VAE
  - Non-linear
  - Probabilistic approach



# Factors for Good Customer Experience

## Casual Learning

- Who to send message to ?
- Small perturbations to messaging policy
- Hidden confounders
- Keeps track of viewership by
  - Push notifications
  - Email reading
  - In-app notifications



# Factors for Good Customer Experience

## Thumbnail Image

- Contextual Bandit
- Personalised image
  - Based on user view history
    - Action
    - Romance
    - Comedy...etc
  - Based on region
  - Based on user's favourite actor/actress


# Softwares and Data Engineering used



- Helps to run queries on big data distributed fashion.
- Either use Hadoop Hive or Hadoop Pig.

cassandra

EVCache

Caching @ Netflix


- They are primary data stores.
- EV Cache is used by Netflix.

NETFLIX


- Hermes is a inhouse Netflix software which delivers data to users in near real time.


- AWS is the software used by both Netflix and Amazon

# Data Used for the System

fractaloo

NETFLIX

| User Behaviour Data (historical data) | Particular Movie/Show Details | Contextual information |
|---|---|---|
| • Shows paused, rewound, fast forwarded, repeated by user<br>• Series not completed or partially completed like a few episodes<br>• Time taken to complete series and reviews given<br>• Similar people likes and dislikes | • Movies/Shows similar to one another in ways of cast, genre, content, release year etc<br>• Movie Credentials like director, cast and crew<br>• Movies description | • Time and date user opened the Netflix<br>• User Profile information as Age, Gender, location, selected content.<br>• the devices user is watching Netflix on |

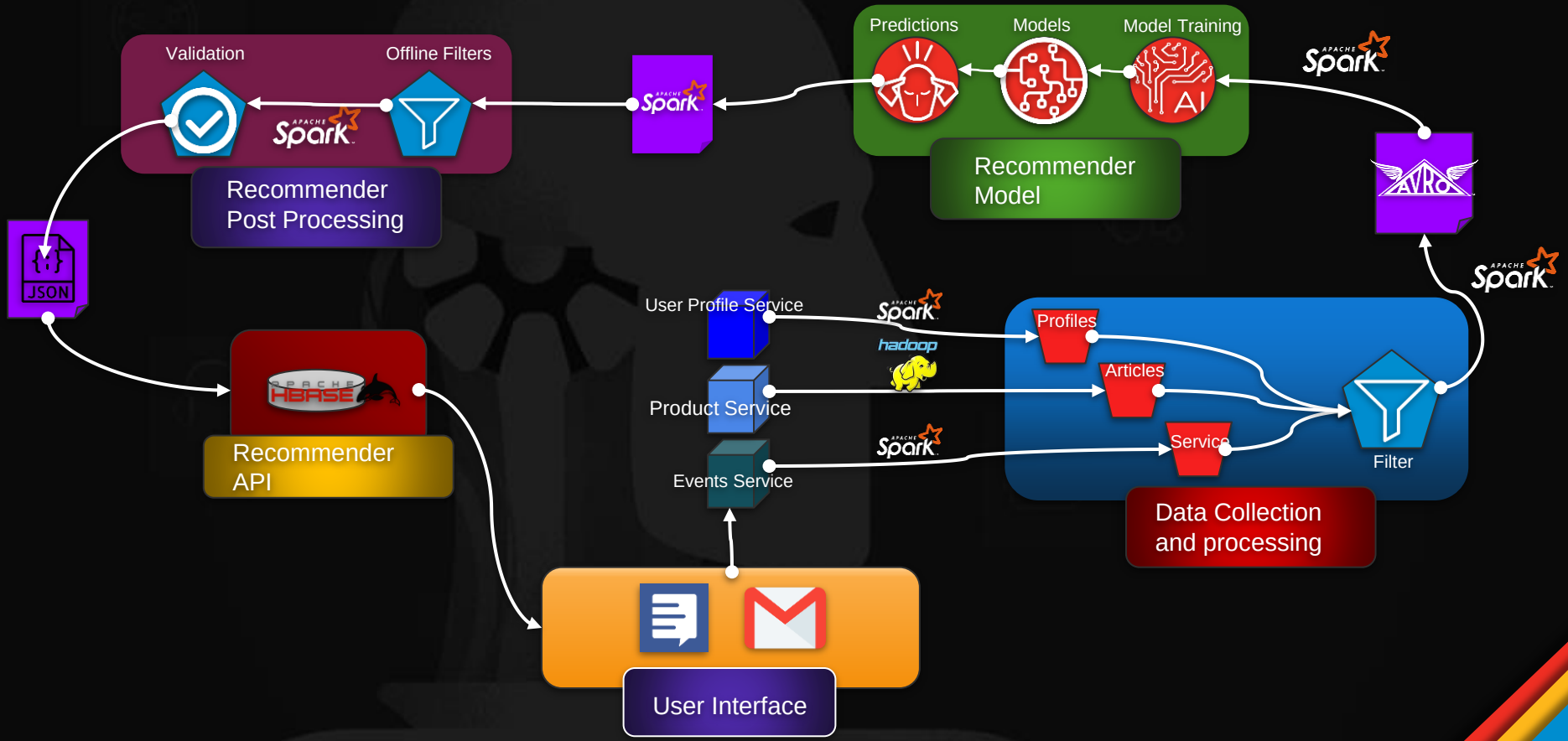
# Data Used for the System

fractal

amazon

| User Behaviour Data (historical data) | Particular Item Details | Contextual information |
|---|---|---|
| • Log on-site activity: clicks, searches, page, and item views, reviews, product buying and cart history<br>• Off-site activities: tracking clicks in emails, in mobile applications, and in their push notifications | • Item title, category, price, description and style<br>• Sales rank for a particular product<br>• Item to Item relationship (such as for laptop we get bag and mouse)<br>• Product Images and CNN features | • User profile data like Age, Gender, Products they like (Such as chocolates, dresses etc)<br>• User's location, Seasonal and time data (In winter no one looks to buy cooler)<br>• Device used<br>• Referral URL |

# Modelling Problem



# Thank You

A presentation by Team Spy AI-Gents


1. Abhishek Bisen
2. Aisha Sartaj
3. Samruddhi Santosh Chilatre
4. Jayesh Kandpal
5. Akhil Chitta
6. Kiran Mahesh Avvari