Master Defence

Privacy Preserving Recommendation systems

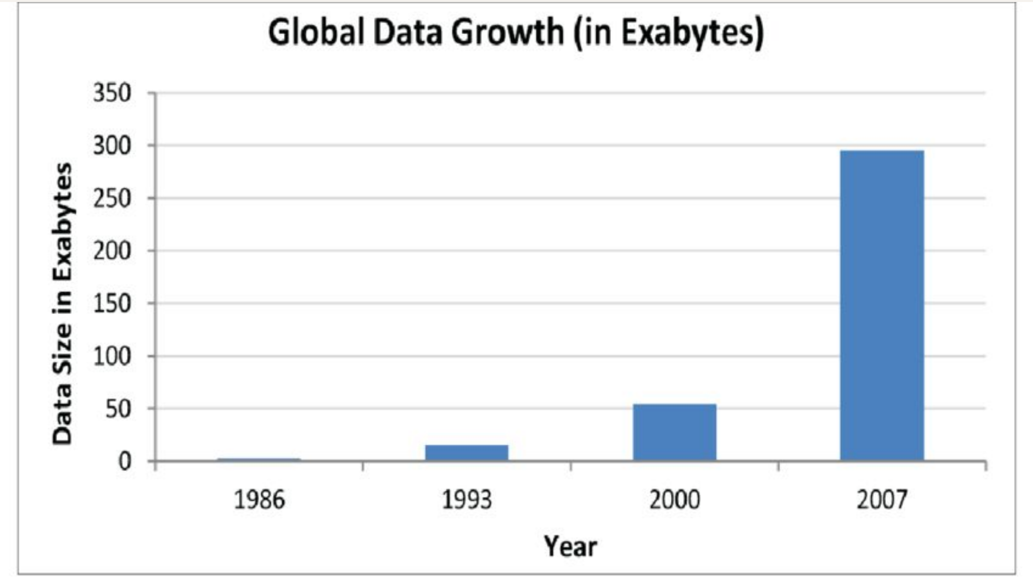
By: Mohamed Naas

Supervisor: Alaa Eddine Belfedhal

Table of contents

01
Introduction

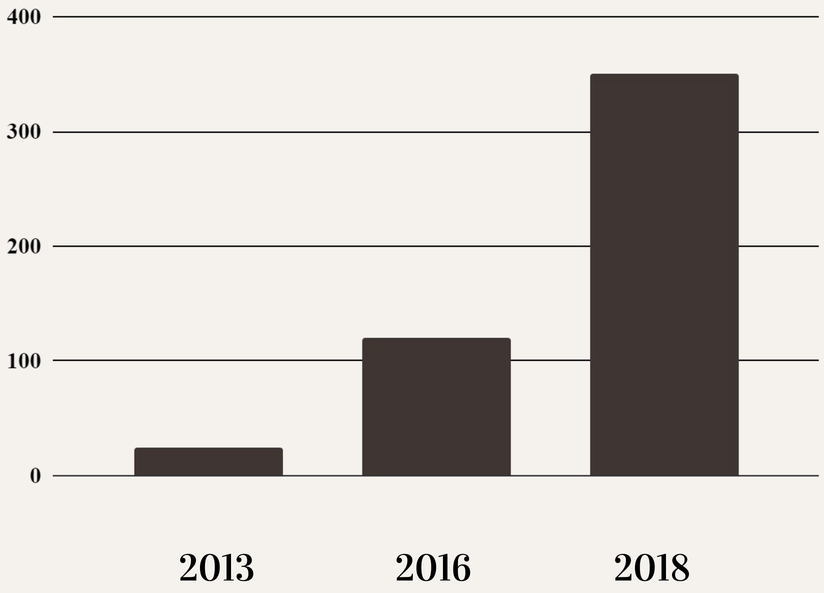
Data Growth



Scientific big data and Digital Earth - Huadong Guo

Incomes of Netflix

Stocks of the streaming company Netflix in dollars before and after 2015 when they build their business model about an optimized recommender system

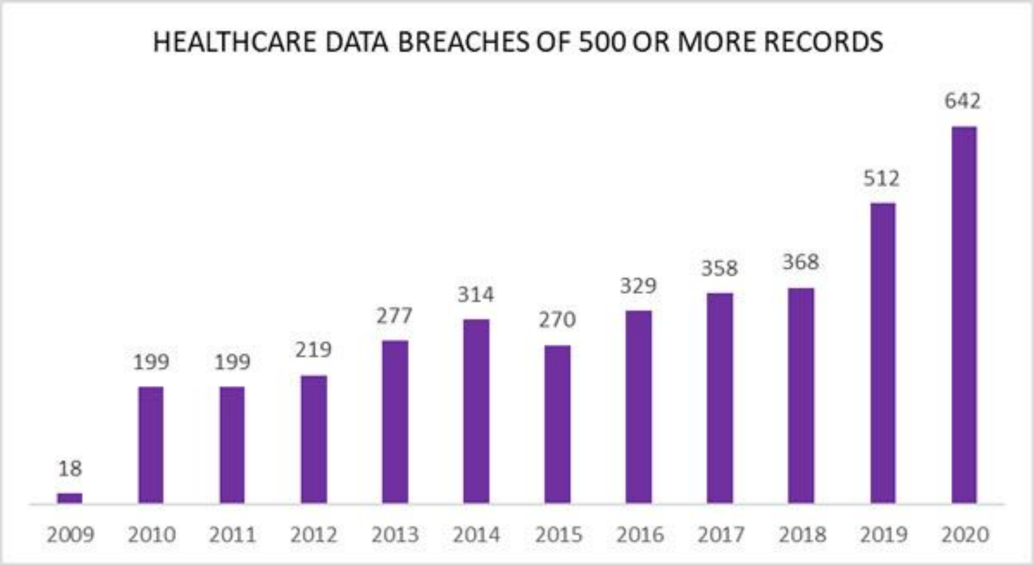


Netflix.com

Every business should have a recommendation system

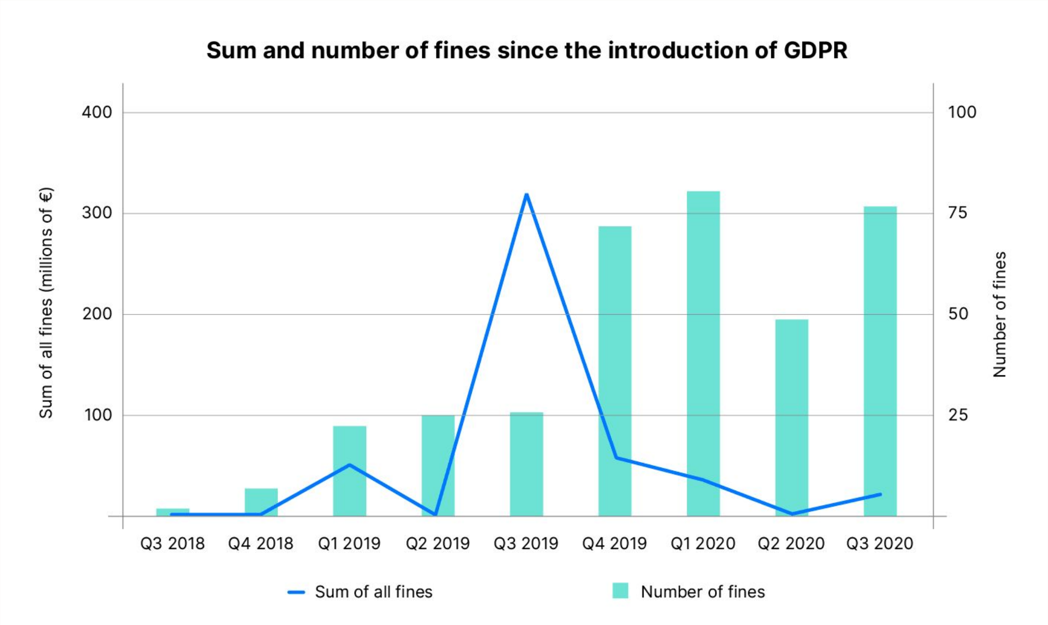
But...

Data Breaches



Hipaa journal 2021

Data Breaches



enforcementtracker.com, provided by CMS Law.Tax

02

Recommendation Systems

Definition

Recommender systems have the goal of generating meaningful recommendations/suggestions to a set of **users** for **items** that might interest them.

Items like movies, music, courses, friends, restaurants, books ...

Again, Why?

Users get what they like

More Watch time

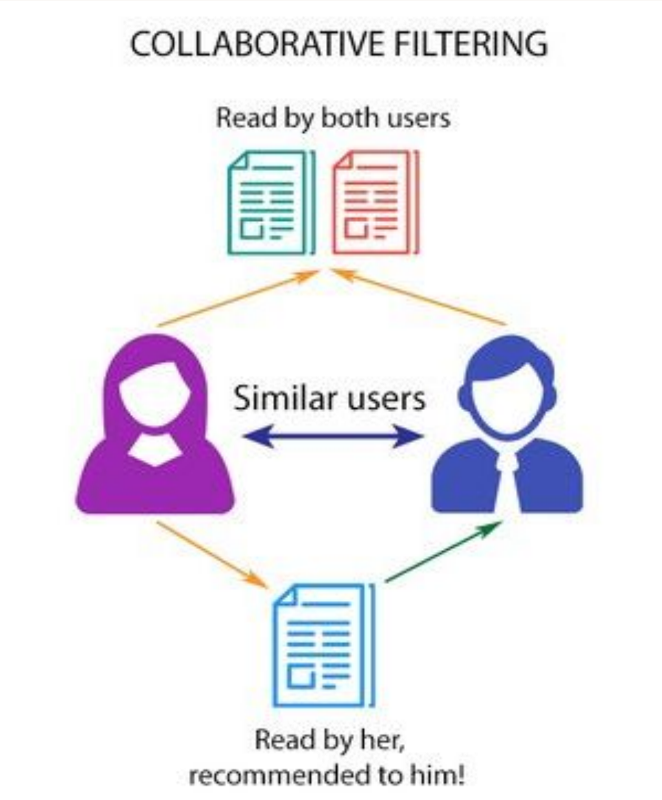
Business gains more money

Approaches

- **Collaborative filtering**
- **Content based recommender systems**
- Knowledge-based recommender system
- Demographic recommender systems
- Hybrid recommender systems

Collaborative filtering

Collaborative filtering (CF) simply-put is recommending items based on the user's previous ratings and on what other similar users liked in the past.



https://www.themarketingtechnologist.co/building-a-recommendation-engine-for-geeksetting-up-the-prerequisites-13

Memory-based Collaborative filtering

Memory-based algorithms use the entire (user-item) data set to generate predictions.

The recommendations are generated on the basis of their neighborhoods.

Memory based CF have two types: user based and item based

Model-based Collaborative filtering

Model-based algorithms build and learn a model from a dataset of ratings and then use that model to generate recommendations in the future.

Model based usually rely on supervised learning or unsupervised learning methods.

Matrix factorization

Matrix factorization is a model based CF approach that generates users and items latent factors (features) from the (item-user) ratings.

$$\overset{\text{Rating Matrix}}{\begin{bmatrix} 4 & & 3 & \\ & 2 & & 1 \\ 5 & 1 & & \\ 2 & & 3 & 2 \end{bmatrix}} = \overset{\text{Users's Latent Factots}}{\begin{bmatrix} 2.2 & 1.25 \\ 2.0 & 2.9 \\ 4.9 & 5.0 \\ 3.8 & 3.2 \end{bmatrix}} \times \overset{\text{Items's Latent Factots}}{\begin{bmatrix} 1.2 & 5.1 & 3.55 & 4.42 \\ 3.8 & 2.4 & 5.2 & 0.9 \end{bmatrix}}$$

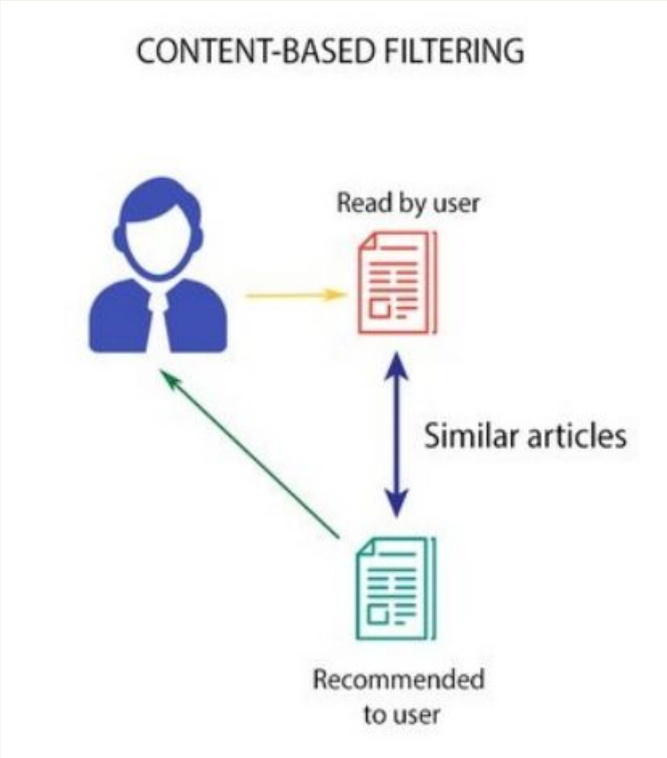
Matrix factorization

Optimization methods:

- Stochastic Gradient Descent
- Alternating Least Squares

Content based recommender systems

Content based algorithms generate recommendations based on the items features and descriptions.



https://www.themarketingtechnologist.co/building-a-recommendation-engine-for-geeksetting-up-the-prerequisites-13

03

Privacy Preserving Machine Learning

PPML?

Privacy preserving machine learning are a set of techniques proposed by the academia to build more privacy friendly models.

Private information?

Gender

Salary

Age

Health status

Location

. . .

Differential Privacy

Differential privacy (DP) constitutes a **standard** to guarantees privacy in statistical analysis and machine learning

Definition: A randomized function K satisfies ϵ-differential privacy if for all data sets D1 and D2 differing in one element (all the possible outputs of k are called S) :

$$\log \frac{\mathbb{P}(M(D) \in S)}{\mathbb{P}(M(D') \in S)} \leq \epsilon$$

Where **ϵ** is the privacy budget.

Differential Privacy

- Local Differential Privacy
- Global Differential Privacy

Homomorphic Encryption

Homomorphic Encryption (HE) is a form of encryption where you can do operation on the encrypted data without the need to decrypt it.

Definition: An encryption scheme is called homomorphic over an operation ” $\star$ ” if it supports the following equation:

$$E(m_1) \star E(m_2) = E(m_1 \star m_2), \quad \forall m_1, m_2 \in M$$

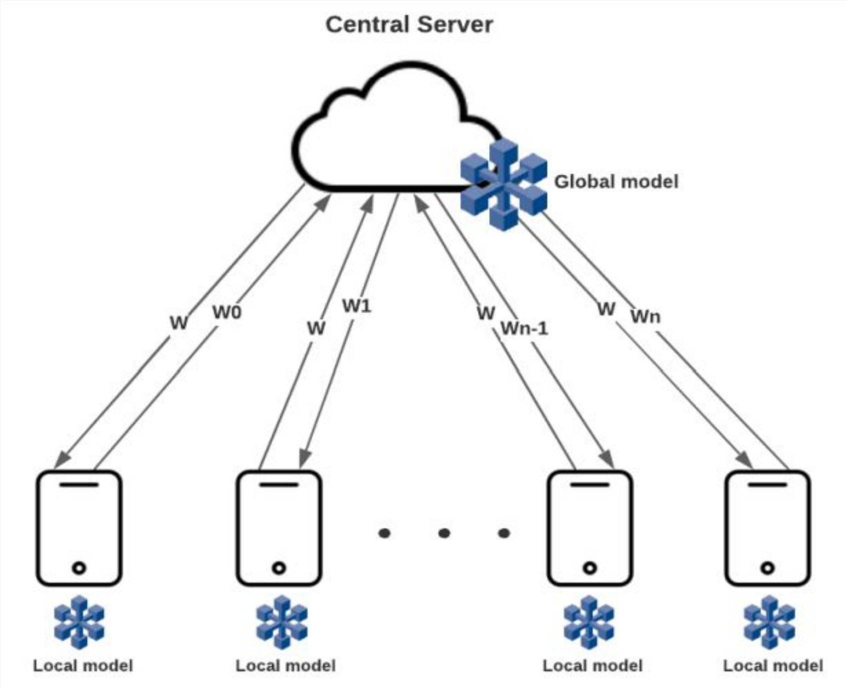
where E is the encryption algorithm and M is the set of all possible messages.

Federated Learning

Federated Learning (FL) enables multiple parties to train a machine learning models without exchanging their local data.

FL usually operate by sending copies of the model to the participating parties where they train the model on their local dataset

Federated Learning



Cross-device federated learning system

Secure Multi Party Computation

Secure Multi Party Computation (SMPC) is a protocol that allows multiple parties to collectively evaluate a function while keeping the inputs of each individual private and without using a trusted third party.

The participants only learn their final result, but not the input data of others.

04

State of Art

Approaches for Privacy Preserving Recommender Systems

- Differential Privacy Solutions
- Homomorphic Encryption Solutions
- Federated Learning Solutions

Differential Privacy Approaches

Paper	Stages	Dimensionality Reduction	Architecture	Differential Privacy Type
Xiao Liu et al	2	No	KNN	Global
Jiang et al	2	No	MF	Local
Shin et al	2	No	MF	Local
Ao Liu et al	1	No	FM + DNN	Condensed
Tao Qi et al	1	No	Neural network	Local

Homomorphic Encryption Solutions

Paper	Architecture	Crypto Service Provider	Phases
Kim et al	Matrix factorization	Yes	3
Badsha et al	KNN	Yes	2
Chai et al	Matrix factorization	No	1

Federated Learning Solutions

Paper	Architecture	Privacy Technique
Ammad et al	Matrix factorization	FL
Chai et al	Matrix factorization	FL + HE
Ying et al	Matrix factorization	FL + Secret Sharing
Tao Qi et al	Neural Network	FL + DP

05

Conclusion

Conclusion

Differential privacy is the most used approach but it effects the recommender performance

Homomorphic encryption does not introduce any additional noise but it is so slow to be practical

Federated learning needs to be combined with other PPML approaches

Significant improvement compared to the previous attempts

The current approaches are not ready to applied for real world application

Building privacy preserving recommenders should be a new research focus.

Thank You