

추천 시스템 소개

이청용 교수

leecy@hansung.ac.kr

- **추천 시스템 개요**

- ✓ 추천 시스템 정의
- ✓ 추천 시스템 도입 배경
- ✓ 추천 시스템의 주요 사례

- **추천 시스템의 역사 및 발전**

- ✓ 1990년대: 추천 시스템의 태동기
- ✓ 2000년대 초반: 협업 필터링과 콘텐츠 기반 필터링의 도입
- ✓ 2010년대: 빅데이터와 머신러닝의 도입
- ✓ 현대의 추천 시스템: 딥러닝과 강화 학습의 발전

- **추천 시스템의 주요 유형**

- ✓ 콘텐츠 기반 필터링
- ✓ 협업 필터링
- ✓ 하이브리드 방법

- **추천 시스템의 핵심 요소**

- ✓ 데이터 수집
- ✓ 사용자 프로파일링
- ✓ 아이템 프로파일링
- ✓ 유사성 측정

- **추천 시스템의 도전과제**

- ✓ 데이터 희소성: 콜드 스타트 문제
- ✓ 확장성: 대규모 데이터 처리의 어려움
- ✓ 편향과 다양성 문제: 필터 버블 현상

추천 시스템 개요

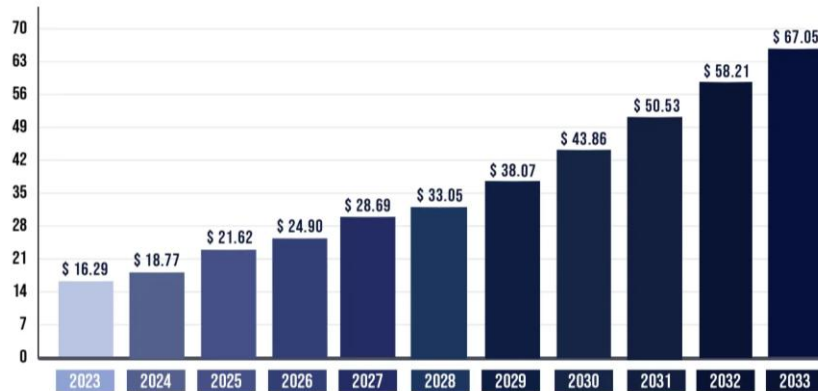
- 추천 시스템이란?

- ✓ 추천 시스템(Recommender System)은 **사용자의 선호도나 과거 행동 데이터**를 기반으로 특정 상품, 콘텐츠, 서비스 등을 **개인 맞춤형으로 예측하고 추천**하는 소프트웨어 시스템
- ✓ 정보가 넘쳐나는 현대 사회에서 사용자가 필요로 하는 정보를 효율적으로 제공함으로써 **정보 과부하 문제를 해결**하고 **사용자 경험을 향상**시키는 것이 추천 시스템의 주요 목표



• 정보 과부화 문제

- ✓ 1990년대 후반 인터넷과 전자 상거래의 급격한 발전으로 인해 사용자는 방대한 양의 정보에 직면
- ✓ 너무 많은 선택지로 인해 사용자는 원하는 상품이나 콘텐츠를 찾기 어려워졌고 이는 구매 또는 소비 과정에서 장애물로 작용
- ✓ 이 문제를 해결하기 위해 추천 시스템이 등장하였으며 사용자에게 의미 있는 정보만을 제공하여 시간과 노력을 절약하는 것을 목표로 설정



E-commerce Market Size and Forecast
source: precedenceresearch.com

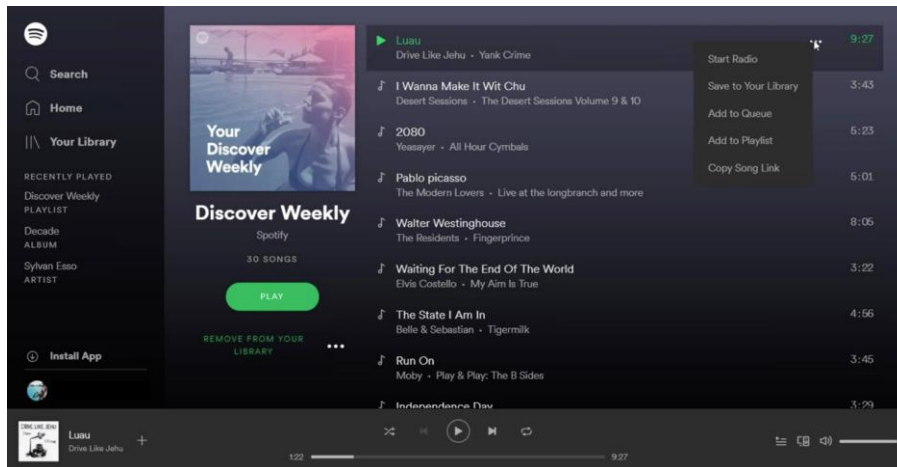


Asia Pacific E-commerce Market Size and Growth 2024 to 2033
source: precedenceresearch.com

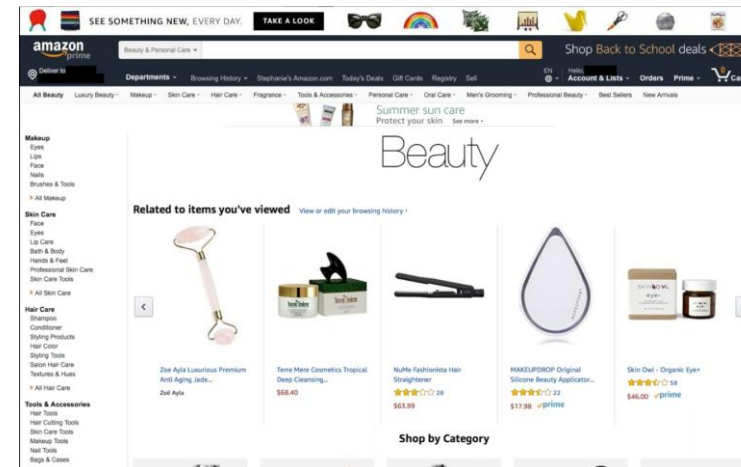
추천 시스템 도입 배경

• 개인화된 경험의 요구

- ✓ 인터넷 사용자의 증가와 함께 **사용자 경험(UX)의 중요성**이 대두
- ✓ 특히 **디지털 서비스**는 개인화된 맞춤형 경험을 제공해야 한다는 필요성 강조
- ✓ 사용자의 선호와 행동을 반영한 개인화된 추천을 제공함으로써 **사용자 만족도와 서비스의 충성도** 제고



source: nngroup.com



source: nngroup.com

- 전자 상거래 및 디지털 마케팅의 발전

- ✓ 전자 상거래의 성장과 함께 사용자에게 적합한 상품을 제시하는 것은 **매출 증가와 직결**
- ✓ 아마존과 같은 대형 전자 상거래 업체들은 **사용자 데이터(구매 이력, 장바구니 등)**를 기반으로 한 추천 시스템을 활용하여 판매를 극대화
- ✓ 기업은 추천 시스템을 통해 사용자 맞춤형 상품을 제시함으로써 **전환율(Conversion Rate)**을 높이고 **구매 유도**를 강화

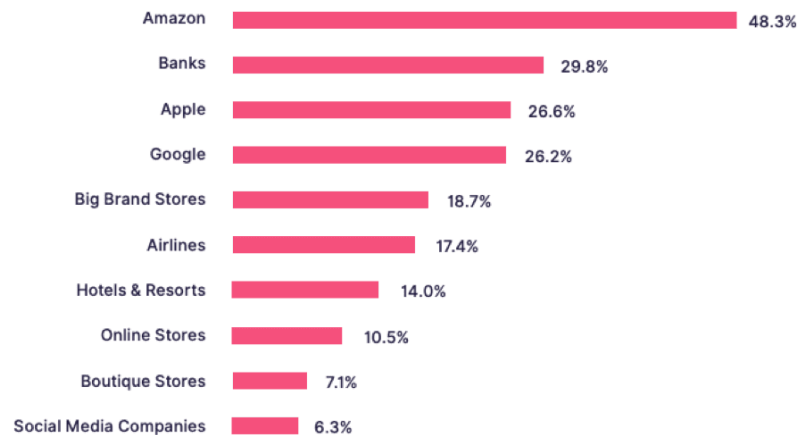
insights

**Amazon's
recommendation
algorithm drives 35% of
its sales**

July 3, 2020

source: evdelo.com/

Which of the following industries or companies do you feel confident your data is being used responsibly?



source: smartinsights.com

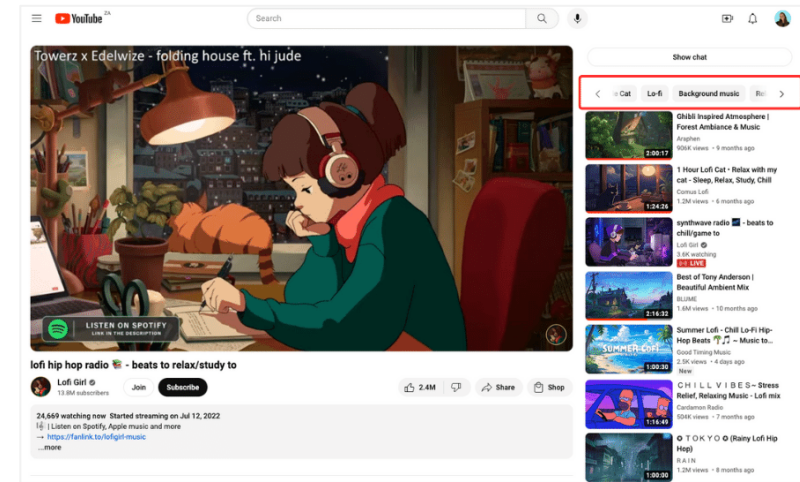
추천 시스템 도입 배경

• 콘텐츠 소비의 변화

- ✓ 미디어 소비 방식이 변화하면서 유튜브 등 플랫폼에서는 방대한 콘텐츠 중 사용자가 선호하는 영상을 개인화하여 추천하는 것이 필수적
- ✓ 사용자가 직접 콘텐츠를 찾기보다는 추천 시스템을 통해 흥미로운 콘텐츠를 탐색
- ✓ 엔터테인먼트와 미디어 스트리밍 산업의 경쟁이 심화되면서 사용자 맞춤형 추천은 사용자의 지속적인 이용과 충성도를 이끄는 핵심 요소로 부상



source: protectdemocracy.org



source: buffer.com

추천 시스템 도입 배경

- 데이터의 축적 및 분석 기술의 발전

- ✓ 추천 시스템이 발전할 수 있었던 또 다른 배경은 **대규모 데이터를 처리하고 분석할 수 있는 기술 발전**
- ✓ 초기의 단순한 필터링 방식에서 벗어나 머신러닝과 딥러닝과 같은 **고도화된 기술을 활용**하여 더욱 정교한 추천이 가능
- ✓ 행동 패턴을 보다 깊이 분석하여 과거에는 파악하기 어려웠던 **사용자의 잠재적인 선호**까지도 반영하는 추천이 가능



source: alleo.tech



source: redbubble.com

• 넷플릭스 (Netflix)

- ✓ 추천 알고리즘: 넷플릭스는 다양한 알고리즘을 결합하여 사용자의 시청 이력, 좋아요/싫어요 반응, 평점 평가, 검색 기록, 시청 시간 등을 분석하여 사용자가 좋아할 만한 콘텐츠를 예측
 - 협업 필터링: 비슷한 취향을 가진 다른 사용자들이 시청한 콘텐츠를 기반으로 추천
 - 콘텐츠 기반 필터링: 사용자가 시청한 콘텐츠의 장르, 감독, 배우, 줄거리 등의 속성을 분석하여 유사한 다른 콘텐츠를 추천
 - Top-N 추천: 사용자의 시청 기록 중에서 가장 자주 시청한 장르나 주제에 맞는 콘텐츠를 N개의 순위로 추천 목록에 제공

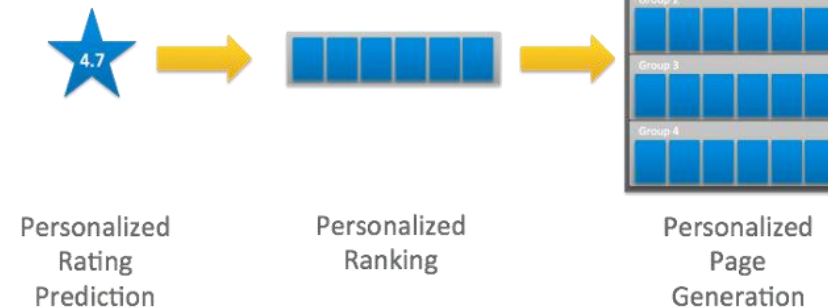
Everything is a Recommendation



Recommendations are driven by machine learning algorithms

Over 80% of what members watch comes from our recommendations

source: mikescogs20.medium.com

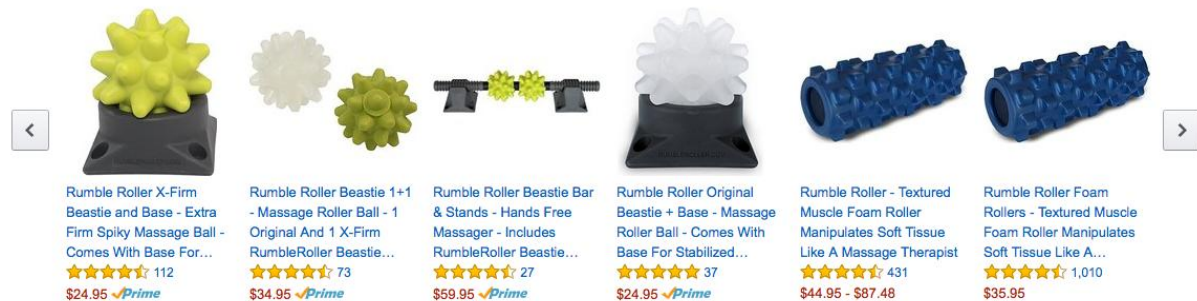


source: mikescogs20.medium.com

• 아마존 (Amazon)

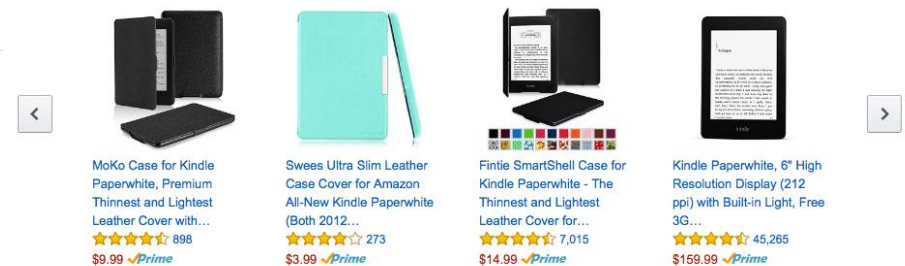
- ✓ 추천 알고리즘: 아마존은 구매 이력, 장바구니에 담은 상품, 검색한 상품, 리뷰와 평점 등의 데이터를 종합하여 개인 맞춤형 상품을 추천하며 주로 아이템 기반 협업 필터링을 사용
 - **아이템 기반 협업 필터링:** 비슷한 상품을 구매한 다른 고객의 행동을 기반으로 추천. 예를 들어, 사용자가 스마트폰을 구매한 후 비슷한 스마트폰 액세서리(케이스, 충전기 등)를 추천 받는 방식
 - **연관 상품 추천:** 특정 상품을 구매한 사용자들이 **함께 구매한 상품**을 추천. 예를 들어, "이 상품을 구매한 고객들은 다음 상품도 구매했습니다."라는 문구와 함께 관련 상품을 제안
 - **타임리밋 추천:** 장바구니에 오랫동안 담아두거나 **자주 검색하는 상품**에 대해 시간 한정 할인이나 프로모션 상품을 추천하여 구매를 유도

Customers Who Bought This Item Also Bought



source: rejoiner.com

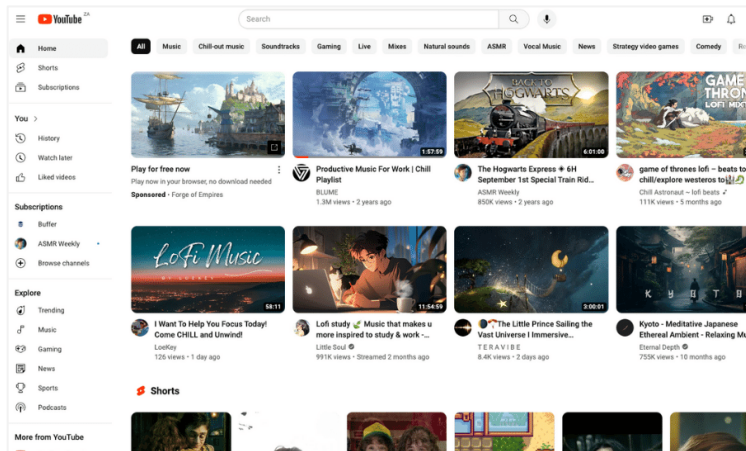
Recommended for You Based on Kindle Paperwhite, 6" High Resolution Display w...



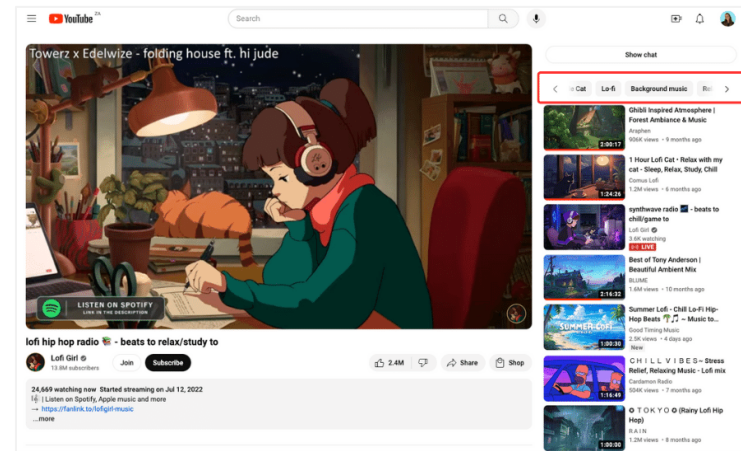
source: rejoiner.com

• 유튜브 (YouTube)

- ✓ 추천 알고리즘: 유튜브는 사용자의 시청 이력, 동영상 시청 시간, 좋아요/싫어요 반응, 댓글 참여도, 구독 채널 등의 다양한 데이터를 활용하여 맞춤형 동영상을 추천
- 협업 필터링: 유사한 취향을 가진 사용자가 시청한 동영상을 추천하여 비슷한 콘텐츠를 발견
- 콘텐츠 기반 필터링: 사용자가 시청한 동영상의 주제, 키워드, 제목, 태그 등을 분석하여 관련된 주제의 동영상을 추천
- Top-N 추천: 사용자가 가장 많이 본 콘텐츠 유형(예: 음악 비디오, 게임, 뉴스)에 따라 최적화된 동영상을 추천 목록에 노출
- 실시간 추천: 사용자의 시청 패턴을 실시간으로 분석하여 시청 도중에 흥미를 잃을 경우 다른 관련 동영상을 빠르게 추천



source: buffer.com

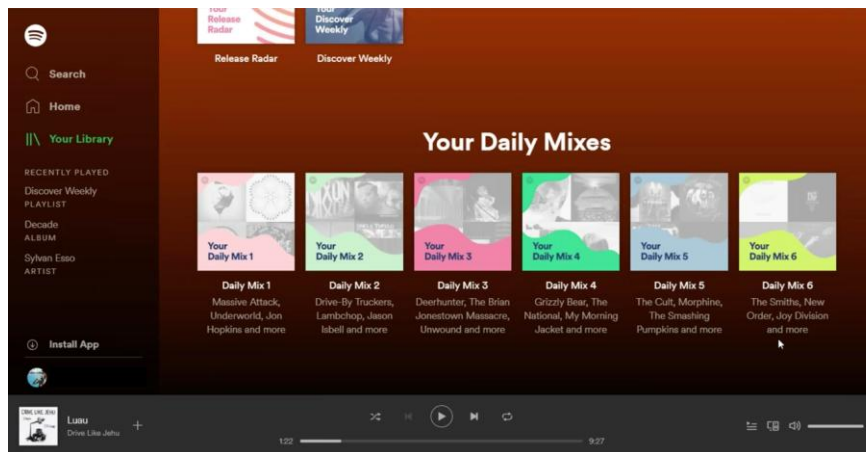


source: buffer.com

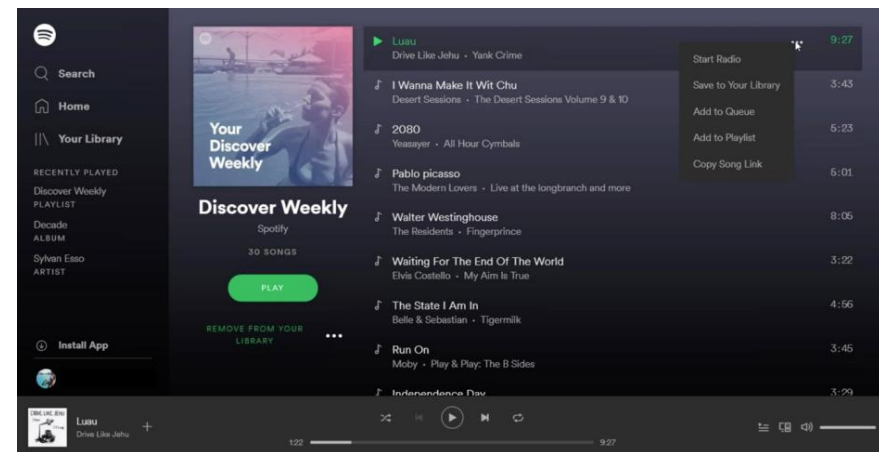
추천 시스템의 주요 사례

• 스포티파이 (Spotify)

- ✓ 추천 알고리즘: 스포티파이는 사용자의 음악 청취 이력, 좋아요/싫어요 클릭, 반복 재생한 곡, 저장한 플레이리스트 등을 분석하여 개인화된 음악을 추천
 - 콘텐츠 기반 필터링: 사용자가 자주 듣는 노래의 장르, 아티스트, 곡의 분위기 등을 분석하여 유사한 음악을 추천
 - 협업 필터링: 다른 사용자들의 청취 패턴을 기반으로 유사한 음악 취향을 가진 사용자가 자주 듣는 곡을 추천
 - Discover Weekly & Release Radar: 매주 스포티파이는 사용자가 자주 듣는 아티스트와 유사한 새로운 음악을 제공하는 Discover Weekly와 사용자의 선호에 맞춘 최신 릴리스 곡을 추천하는 Release Radar 기능을 제공
 - Daily Mix: 사용자의 청취 이력에 따라 매일 새로운 플레이리스트를 생성하여 개인화된 음악 추천을 제공



source: rejoiner.com



source: rejoiner.com

추천 시스템의 역사 및 발전

- 추천 시스템의 개념은 1990년대 중반에 등장했으며 주로 사용자가 제공하는 명시적 피드백(예: 평점, 리뷰)을 활용하여 간단한 규칙에 따라 추천을 제공
 - ✓ **명시적 피드백 기반:** 사용자가 직접 제공한 **평점이나 리뷰**와 같은 피드백을 기반으로 추천 제공. 이때의 추천 시스템은 사용자의 과거 행동을 분석하거나 명시적인 선호 데이터를 바탕으로 추천을 생성
 - ✓ **규칙 기반 추천:** 간단한 규칙을 기반으로 추천이 이루어졌으며, 특정 상품이나 콘텐츠에 대해 **일정한 점수 이상의 평점**을 부여한 사용자에게 **비슷한 아이템을 추천**하는 방식을 주로 사용
 - ✓ **사례:** 넷스케이프(Netscape) 같은 초창기 인터넷 브라우저에서도 기본적인 추천 기능이 도입되었으나 사용자 데이터를 심층적으로 활용하는 데 한계가 존재

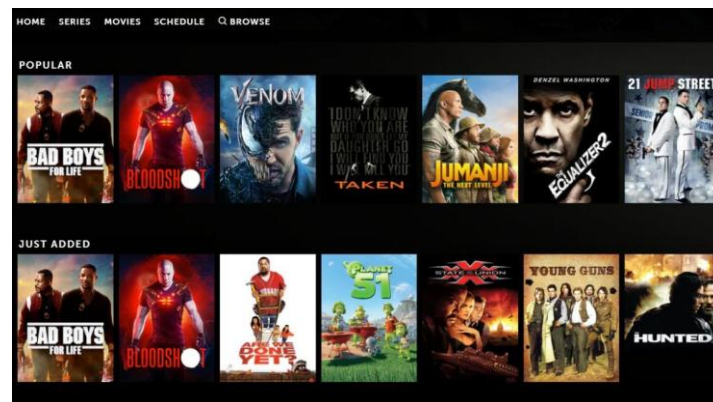
2000년대 초반: 협업 필터링과 콘텐츠 기반 필터링의 도입

- 2000년대에 들어서면서 추천 시스템은 협업 필터링과 콘텐츠 기반 필터링이라는 두 가지 주요 방식으로 발전. 사용자 데이터를 보다 복잡하게 분석하여 더 정확하고 개인화된 추천 제공의 가능성을 마련
 - ✓ **협업 필터링**: 여러 사용자의 행동 데이터를 바탕으로 유사한 사용자가 선호하는 아이템을 추천. 이는 사용자가 제공하는 명시적 피드백뿐만 아니라 다른 사용자와의 상관관계를 통해 추천 제공
 - **사례**: 아마존은 협업 필터링을 통해 "이 상품을 구매한 고객은 이러한 상품도 구매했습니다."와 같은 방식으로 맞춤형 상품을 추천하기 시작
 - ✓ **콘텐츠 기반 필터링**: 사용자가 과거에 선호한 아이템의 특성을 분석하고, 이와 유사한 특성을 가진 새로운 아이템을 추천. 주로 **아이템의 메타데이터**(예: 영화의 장르, 감독, 배우, 책의 저자 등)를 분석하여 유사한 콘텐츠를 추천
 - **사례**: 영화 추천 서비스에서는 사용자가 선호한 영화의 장르, 감독, 배우 등의 정보를 기반으로 유사한 영화를 추천하기 시작



source:

Linden, G., Smith, B., & York, J. (2003). Amazon. com recommendations: Item-to-item collaborative filtering. *IEEE Internet computing*, 7(1), 76-80.



source: geekycodesin.wordpress.com

2010년대: 빅데이터와 머신러닝의 도입

- 2010년대에 들어서면서 추천 시스템은 빅데이터와 머신러닝 기술을 적극 활용. 사용자의 행동 데이터를 실시간으로 수집, 처리, 분석할 수 있는 기술이 발전하면서 추천 시스템의 정확도 크게 향상
 - ✓ **대규모 데이터 처리**: 사용자의 행동 데이터가 폭발적으로 증가하면서 추천 시스템은 더 많은 데이터를 실시간으로 처리하기 위해 **분산 컴퓨팅 기술**(예: 아파치 하둡, 아파치 스파크)이 도입되었으며 대규모 데이터를 처리하는 데 적합한 시스템이 구축
 - ✓ **머신러닝 기반 추천**: 머신러닝 알고리즘을 활용하여 사용자 행동 패턴을 더욱 **정교하게 분석 및 예측** 가능. 이는 단순한 협업 필터링과 콘텐츠 기반 필터링을 넘어 복잡한 데이터 패턴을 찾아내는 데 도움을 제공
 - **행렬 분해**(Matrix Factorization), **잠재 요인 모델링**(Latent Factor Modeling)과 같은 기법이 도입되어 사용자의 잠재적 선호도를 정확하게 예측 가능
 - **사례**: 넷플릭스는 2006년부터 시작된 Netflix Prize 대회를 통해 행렬 분해 알고리즘을 활용하여 추천 시스템의 성능을 크게 개선



Rank	Team Name	Best Test Score	% Improvement	Best Submit Time
Grand Prize - RMSE = 0.8567 - Winning Team: BellKor's Pragmatic Chaos				
1	BellKor's Pragmatic Chaos	0.8567	10.06	2009-07-26 18:18:28
2	The Ensemble	0.8567	10.06	2009-07-26 18:38:22
3	Grand Prize Team	0.8582	9.90	2009-07-10 21:24:40
4	Opera Solutions and Vandelay United	0.8588	9.84	2009-07-10 01:12:31
5	Vandelay Industries I	0.8591	9.81	2009-07-10 00:32:20
6	Pragmatic Theory	0.8594	9.77	2009-06-24 12:06:56
7	BellKor in BioChaos	0.8601	9.70	2009-05-13 08:14:09
8	Dace	0.8612	9.59	2009-07-24 17:18:43

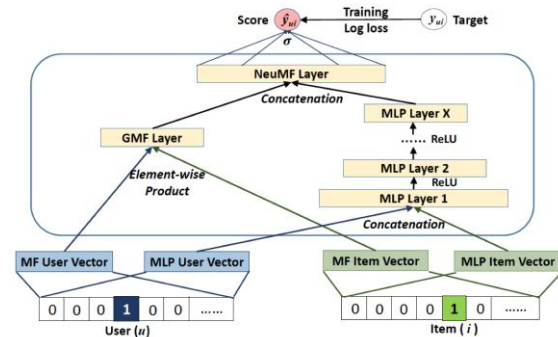
source: wired.com



source:

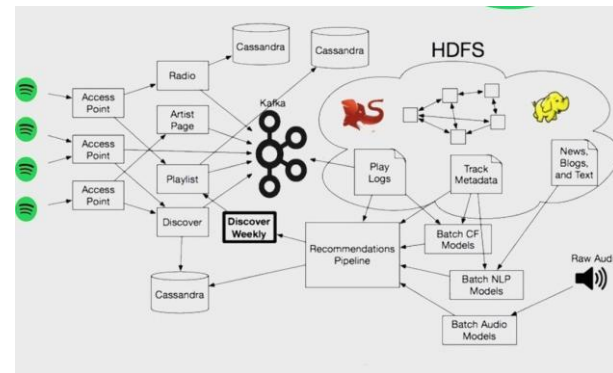
Koren, Y., Bell, R., & Volinsky, C. (2009). Matrix factorization techniques for recommender systems. *Computer*, 42(8), 30-37.

- 2020년대에 접어들면서 추천 시스템은 딥러닝 및 강화학습과 같은 첨단 기술을 활용하여 더 높은 수준으로 발전
 - ✓ **딥러닝을 활용한 추천 시스템:** 딥러닝 모델은 방대한 양의 데이터를 처리하고 사용자와 아이템 간의 **복잡한 상호작용**을 더 정밀하게 학습 가능. 특히 신경망 기반 추천 시스템(Neural Collaborative Filtering, NCF)이 등장하여 사용자와 아이템 간의 **비선형 관계를 학습**하고 추천 정확도를 향상
 - **사례:** 유튜브는 딥러닝을 활용하여 사용자의 시청 이력을 기반으로 가장 관련성 높은 동영상을 추천하며 콘텐츠의 썸네일을 개인 맞춤형으로 제공하여 더 높은 클릭률을 유도
 - ✓ **강화 학습을 활용한 실시간 추천:** 강화 학습은 사용자가 추천된 아이템에 어떻게 반응하는지 실시간으로 피드백을 받아 시스템이 스스로 학습하고 최적의 추천을 제공하는 방법. 이를 통해 시스템은 사용자의 **변화하는 선호도**를 반영하여 **실시간으로 추천**을 개선 가능
 - **사례:** 전자 상거래와 미디어 플랫폼에서 강화 학습을 활용하여 사용자의 실시간 행동을 기반으로 즉각적으로 반응하는 맞춤형 추천을 제공. 예를 들어, 스포티파이는 사용자의 청취 패턴을 실시간으로 분석하여 즉각적으로 맞춤형 곡이나 플레이리스트를 추천



source:

He, X., Liao, L., Zhang, H., Nie, L., Hu, X., & Chua, T. S. (2017, April). Neural collaborative filtering. In *Proceedings of the 26th international conference on world wide web* (pp. 173-182).



source: medium.com

추천 시스템의 주요 유형

- 아이템의 콘텐츠에 집중하여 사용자가 선호하는 **아이템의 특성을 분석**한 후 유사한 특성을 가진 다른 아이템을 추천하는 방식
 - ✓ 작동 원리
 - 사용자가 선호한 **아이템의 속성**(예: 영화의 장르, 배우, 감독, 출판물의 주제 등)을 분석
 - 사용자의 이전 행동이나 명시적 평가를 통해 선호도를 파악한 후 그와 **유사한 속성을 가진 다른 아이템을 추천**
 - ✓ 장점
 - 사용자에게 대한 데이터가 충분하지 않아도 아이템의 속성만으로 신규 아이템에 대한 추천이 가능하므로 **콜드 스타트 문제**에서 **상대적으로 유리**
 - 사용자의 명시적 선호도(좋아하는 장르, 스타일 등)를 기반으로 직관적인 추천이 가능
 - ✓ 단점
 - 사용자가 선호하는 아이템의 범위에 갇혀 **새로운 경험**을 제공하는 데 **한계**가 존재
 - 아이템 속성을 수집하고 처리할 때 **시간이 많이 소요**될 수 있으며 아이템의 특성을 정의하기 어렵거나 복잡하면 **성능이 저하**될 가능성이 존재

- 사용자 기반 협업 필터링 (User-based Collaborative Filtering)

- ✓ 작동 원리

- 사용자의 선호도를 다른 사용자와 비교하여 나와 **비슷한 취향**을 가진 사용자를 탐색.
 - 비슷한 취향을 가진 다른 사용자가 좋아한 아이템을 나에게 추천. 즉, '비슷한 사람들'이 선호한 아이템이 나에게 적합할 가능성이 높다는 가정에 기반

- ✓ 장점

- **다양한 추천**을 제공할 수 있어 사용자가 좋아할 가능성이 높은 **새로운 아이템**을 발견하여 제공
 - 아이템의 특성을 직접 분석할 필요가 없이 사용자 행동만으로 추천이 가능

- ✓ 단점

- **콜드 스타트 문제**: 신규 사용자나 아이템에 대한 데이터가 부족할 경우 유사성을 기반으로 한 추천이 어려움 존재
 - **확장성 문제**: 많은 사용자와 아이템이 있는 경우 계산 복잡도가 증가하여 실시간 추천에 어려움이 존재

- **아이템 기반 협업 필터링 (Item-based Collaborative Filtering)**

- ✓ **작동 원리**

- 사용자가 선호한 아이템의 특성을 분석하고 그 아이템과 유사한 다른 아이템 추천
 - **아이템 간의 유사성**을 측정한 후 사용자가 선호하는 아이템과 **비슷한 속성**을 가진 **아이템 추천**

- ✓ **장점**

- 사용자 행동에 따라 직관적인 추천이 가능하며 사용자 간의 데이터가 부족해도 **아이템의 유사성만으로 추천**이 가능
 - 상대적으로 계산 복잡도가 낮아 **대규모 데이터**에서도 **성능이 유지**

- ✓ **단점**

- 아이템 간의 유사성이 명확하지 않을 때 추천 품질이 저하될 가능성 존재
 - **새로운 사용자**에게는 적절한 **추천 어려움** 존재

- 각 알고리즘의 단점을 상호 보완하여 추천 시스템의 성능을 높이는 방식이며, 일반적으로 두 가지 이상의 알고리즘을 결합하거나 각각의 알고리즘이 제안한 추천 결과를 통합하여 최적의 추천을 생성

✓ 작동 원리

- **혼합형 방식:** 콘텐츠 기반 필터링과 협업 필터링을 동시에 사용하여 두 결과를 결합한 후 사용자에게 최종 추천을 제공
- **단계적 방식:** 먼저 협업 필터링을 사용해 추천을 생성하고 이후 콘텐츠 기반 필터링으로 해당 추천을 필터링하는 방식으로 사용

✓ 장점

- **정확성 향상:** 각 알고리즘이 가진 한계를 **상호 보완**할 수 있어 더 정확하고 다각적인 추천이 가능
- **다양성 증대:** 협업 필터링의 추천에 콘텐츠 기반 필터링을 추가함으로써 추천 목록의 **다양성 증가**

✓ 단점

- 두 알고리즘을 결합함에 따라 **계산 복잡도가 증가**하고 시스템 구현과 **유지보수 어려움** 존재
- 데이터가 충분히 **확보되지 않은 경우** 각 알고리즘의 결합이 오히려 **추천 성능 저해**

추천 시스템의 핵심 요소

- 명시적 피드백(Explicit Feedback)

- ✓ 사용자가 의도적으로 남긴 명확한 반응 데이터를 의미. 명시적 피드백은 **사용자가 선호를 직접적으로 표현**하는 방식으로 수집되며 사용자 선호도를 명확하게 반영하는 중요한 정보
 - **평점**: 사용자가 콘텐츠나 상품에 대해 부여하는 점수(예: 1~5점 평점). 넷플릭스에서는 시청자가 영화나 드라마에 대해 별점 평가를 남기면 추천 알고리즘에 반영
 - **리뷰**: 상품이나 콘텐츠에 대해 사용자가 작성한 텍스트 리뷰. 아마존은 상품에 대한 리뷰를 바탕으로 추천 품질을 향상
 - **좋아요/싫어요**: 유튜브나 소셜 미디어에서 사용자들이 영상이나 게시물에 대해 남기는 좋아요나 싫어요의 클릭 기록

- 암시적 피드백(Implicit Feedback)

- ✓ 사용자가 직접적으로 선호를 표현하지는 않았지만 행동을 통해 **간접적**으로 나타나는 데이터. 암시적 피드백은 사용자의 행동 패턴을 통해 선호도를 추론할 수 있으며 더 많은 양의 데이터를 수집할 수 있다는 장점이 존재
 - **클릭**: 사용자가 웹사이트에서 클릭한 모든 항목의 기록. 예를 들어, 사용자가 특정 상품 페이지를 여러 번 방문했지만 구매하지 않았다면 이는 사용자가 관심을 가지고 있음을 의미
 - **조회 시간**: 사용자가 콘텐츠를 얼마나 오래 시청하거나 읽었는지를 기록. 유튜브에서는 동영상을 시청한 시간에 따라 추천 알고리즘이 조정되며 짧은 시간 동안 여러 개의 영상을 본 사용자에게는 빠른 템포의 콘텐츠를 추천
 - **구매 기록**: 전자 상거래에서 사용자가 구매한 상품의 기록. 아마존은 구매 이력을 분석하여 비슷한 상품을 추천
 - **검색 기록(Search History)**: 사용자가 검색창에 입력한 키워드와 그에 따른 검색 결과를 클릭한 이력. 사용자가 검색한 상품 카테고리나 키워드가 추천 시스템에서 중요한 데이터로 활용

- 개별 사용자의 선호와 행동 패턴을 분석하여 사용자의 관심사와 취향을 반영한 프로파일을 만드는 과정. 이를 통해 추천 시스템은 사용자에게 더 정확한 맞춤형 추천을 제공
 - ✓ **행동 기반 프로파일링**: 사용자의 클릭, 시청, 구매 이력 등의 **행동 데이터**를 기반으로 그들의 관심사를 추정. 예를 들어, 스포티파이는 사용자의 음악 청취 패턴을 분석하여 특정 장르나 아티스트에 대한 선호도를 파악
 - ✓ **명시적 선호 프로파일링**: 사용자가 직접 입력한 **선호 정보나 평가 데이터**를 기반으로 프로파일을 생성. 예를 들어, 사용자가 특정 영화에 대해 높은 평점을 남겼다면 비슷한 장르나 감독의 다른 작품도 추천에 반영
 - ✓ **잠재 요인 모델링(Latent Factor Modeling)**: 사용자의 명시적인 평가 데이터나 행동 패턴을 바탕으로 개별적인 선호도와 관련된 잠재적 요소들을 추정. 이는 행렬 분해(Matrix Factorization) 기법 등을 통해 사용자의 숨겨진 선호를 발견하여 추천 품질을 높이는 데 활용

- 아이템 프로파일링은 상품이나 콘텐츠 자체의 속성을 분석하여 추천에 활용하는 과정. 이는 콘텐츠 기반 필터링에서 중요한 역할을 수행
 - ✓ **메타데이터 분석:** 영화, 책, 상품 등의 **메타데이터**(장르, 주제, 출판사, 감독, 배우, 카테고리 등)를 활용하여 아이템의 특성을 파악. 예를 들어, 넷플릭스는 영화의 장르, 감독, 주연 배우 등의 정보를 분석하여 유사한 영화들을 추천.
 - ✓ **텍스트 마이닝(Text Mining):** 리뷰, 아이템 설명 등 **텍스트 데이터를 분석**하여 아이템에 대한 더 구체적인 정보를 추출. 이 과정을 통해 콘텐츠의 주제나 분위기를 파악할 수 있으며 이를 바탕으로 관련 있는 다른 콘텐츠를 추천
 - ✓ **음악 및 이미지 분석:** 스포티파이와 같은 음악 스트리밍 서비스는 음원 파일을 분석하여 곡의 리듬, 템포, 악기 구성 등을 기반으로 유사한 음악을 추천. 이미지 기반의 분석도 쇼핑 플랫폼에서 활용되며 유사한 디자인이나 색상의 제품을 추천하는 데 활용

- 피어슨 상관계수(Pearson Correlation Coefficient)

- ✓ 피어슨 상관계수는 두 변수 간의 선형 관계를 측정하는 방법으로 사용자나 아이템 간의 상관관계를 계산하는 데 자주 사용
 - **작동 원리:** 두 변수의 공분산을 각 변수의 표준편차로 나누어 상관관계를 구하며 -1에서 1 사이의 값으로 표현. 1에 가까울수록 양의 상관관계가 크다는 의미이며 -1에 가까울수록 반대의 경향을 의미
 - **적용:** 사용자 간의 평점 패턴을 분석하거나 아이템의 선호도 간의 상관관계를 측정하는 데 활용

- 코사인 유사도(Cosine Similarity)

- ✓ 코사인 유사도는 두 벡터 사이의 각도를 기반으로 유사성을 측정하는 방법. 주로 사용자와 아이템 간의 **벡터화된 데이터를 비교**하는 데 사용
 - **작동 원리:** 두 벡터의 내적을 각 벡터의 크기로 나누어 계산하며 값이 1에 가까울수록 두 벡터(사용자 또는 아이템)가 유사하다고 판단. 예를 들어, 두 사용자가 비슷한 영화를 시청하고 유사한 평점을 남겼다면 이들의 벡터 간 각도는 작고 코사인 유사도는 높게 도출
 - **적용:** 협업 필터링에서 사용자 혹은 아이템 간의 유사성을 계산하는 데 자주 사용

추천 시스템의 도전과제

- 데이터 희소성은 추천 시스템이 충분한 사용자나 아이템 데이터를 확보하지 못했을 때 발생하는 문제로, 특히 콜드 스타트(Cold Start) 문제와 밀접한 관련성 존재

✓ 신규 사용자 문제

- 신규 사용자가 플랫폼에 처음 가입하면 해당 사용자의 과거 행동 데이터가 부족하기 때문에 개인화된 추천을 제공하는데 현실적 어려움 존재
- 추천 시스템은 사용자의 취향을 파악할 만한 정보가 거의 없기 때문에 기본적인 추천을 제공하거나 모든 사용자에게 동일한 아이템을 추천하는 방식으로 동작

✓ 신규 아이템 문제

- 새로운 상품이나 콘텐츠가 추가될 때 아이템에 대한 사용자 피드백이나 상호작용 데이터가 없기 때문에 추천 시스템이 해당 아이템을 추천하는 데 어려움이 존재하며, 이는 콘텐츠 기반 필터링보다 협업 필터링에서 더 두드러지게 강조

- 추천 시스템은 많은 사용자와 아이템을 처리해야 하므로 데이터의 양이 커질수록 확장성 문제가 발생. 특히, 실시간으로 대규모 데이터를 처리하고 적시에 맞춤형 추천을 제공하는 것은 기술적인 도전과제

- ✓ 데이터 처리의 복잡성

- 사용자와 아이템의 수가 기하급수적으로 늘어날수록 추천을 생성하는 데 필요한 계산 복잡도 증가
- 이를 처리하기 위해서는 대규모 데이터에 대한 실시간 분석과 계산이 필요하며 이는 서버 자원과 알고리즘의 효율성에 부담을 초래

- ✓ 실시간 추천의 어려움

- 사용자 행동이 실시간으로 변화하는 환경에서는 빠른 속도로 데이터를 처리하고 즉각적인 추천을 생성하는 것이 매우 중요.
- 대규모 데이터를 실시간으로 분석하는 것은 많은 연산 자원을 필요로 하며 이는 성능 저하의 원인으로 작동

- 추천 시스템이 특정 콘텐츠나 상품을 지나치게 반복해서 추천할 경우 사용자에게 제한된 경험만을 제공하는 편향(Bias)과 다양성 부족 문제가 발생

- ✓ 편향 문제

- 추천 시스템은 사용자 행동 데이터를 기반으로 작동하기 때문에 사용자가 이전에 선호했던 콘텐츠나 상품에만 집중하는 경향이 존재
- 사용자는 새로운 콘텐츠를 발견할 기회를 잃고 추천 시스템은 특정 유형의 아이템만을 추천하는 편향 발생

- ✓ 다양성 부족과 필터 버블

- 필터 버블은 사용자가 특정 취향에 맞는 콘텐츠만 추천 받고, 그 외의 정보에 접근할 기회를 제한 받는 현상
- 이로 인해 사용자는 자신의 세계관과 선호에 맞는 정보만 접하게 되어 정보의 편중이 발생

감사합니다