

### Data Science Use Case dalam Industri Retail: Netflix

Data science adalah ilmu yang menggabungkan komponen matematika, statistika dan ilmu komputer dengan tujuan analisa data dari kumpulan data baik skala kecil maupun besar dengan mengaplikasikan algoritma tertentu dengan tujuan untuk *data mining* sehingga mendapatkan pola data serta dapat melakukan prediksi dengan cukup akurat yang dapat membantu dalam pengambilan keputusan dan dapat digunakan untuk membuat sistem yang cerdas (AI) dan dapat terus belajar dengan sendirinya (machine learning). (Jimmy, 2019). Artikel ini akan membahas bagaimana Netflix menggunakan Data Science untuk memberikan rekomendasi yang relevan dan menarik bagi penggunanya. Netflix merupakan sebuah *online streaming platform* yang sangat populer digunakan beberapa tahun terakhir. Netflix menggunakan banyak algoritma demi memberikan pengalaman menonton terbaik bagi para penggunanya. Salah satu algoritma yang digunakan Netflix adalah Sistem Rekomendasi untuk memberikan saran konten yang relevan.

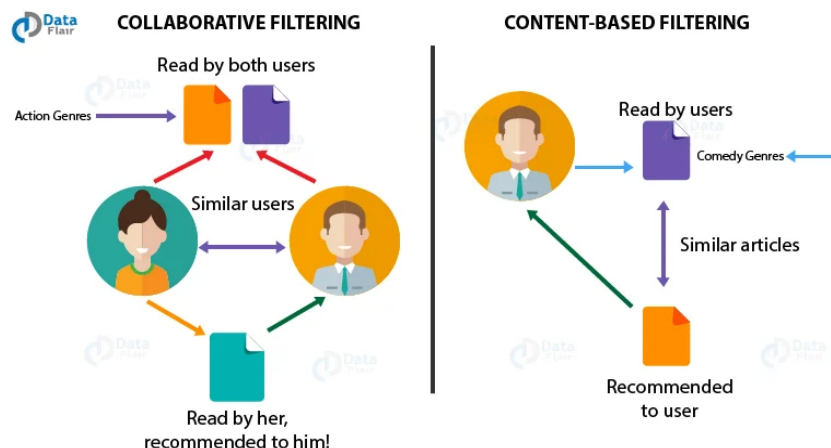
Sistem rekomendasi adalah *platform* yang menyediakan berbagai konten bagi penggunanya berdasarkan preferensi dan kesukaan mereka. Sistem ini mengambil informasi tentang pengguna sebagai *input*. Sistem rekomendasi juga memanfaatkan berbagai algoritma *machine learning*. Di Netflix, sistem rekomendasi mencari film yang mirip dengan yang ditonton atau disukai sebelumnya oleh pengguna. Oleh karena itu, berdasarkan film yang ditonton, Netflix memberikan rekomendasi film yang memiliki tingkat kemiripan yang sama. Terdapat dua jenis sistem rekomendasi:

1. Sistem Rekomendasi Berbasis Konten

Dalam sistem rekomendasi berbasis konten, latar belakang pengetahuan produk dan informasi dari pengguna dipertimbangkan. Contohnya, jika *user* pernah menonton film bergenre komedi, sistem rekomendasi berbasis konten akan memberi saran untuk film serupa yang bergenre sama.

2. Sistem Rekomendasi *Collaborative Filtering*

*Collaborative Filtering* memberikan rekomendasi berdasarkan profil serupa dari penggunanya. Salah satu keuntungan utama dari *collaborative filtering* adalah bahwa hal itu tidak tergantung pada pengetahuan produk. Contohnya, jika profil A menonton genre komedi, kartun dan horor dan profil B menonton genre kartun, horor, dan dokumenter, maka profil A juga akan menyukai dokumenter dan B akan menyukai genre komedi.



Sumber: DataFlair

Ada juga jenis sistem rekomendasi yang menggabungkan dua teknik sistem rekomendasi di atas. Bentuk sistem rekomendasi ini dikenal dengan *Hybrid Recommendation System*. Netflix menjadikan sistem ini sebagai sistem rekomendasi utama untuk merekomendasikan konten kepada penggunanya.

Netflix menggunakan Singular Value Decomposition (SVD) untuk menyediakan embedding dimensi yang optimal bagi penggunanya dan mereka juga menggunakan Restricted Boltzmann Machines (RBM) untuk meningkatkan kemampuan model *collaborative filtering*. Kedua algoritma ini memberikan hasil terbaik. Kombinasi linier dari kedua algoritma ini mengurangi RMSE menjadi 0,88. Namun, setelah pengurangan RMSE dan peningkatan akurasi, Netflix mengalami dua tantangan utama. Pertama, data yang disediakan selama kompetisi terdiri dari 100 juta peringkat film, dibandingkan dengan lebih dari 5 miliar peringkat yang dibuat oleh Netflix. Selain itu, algoritma bersifat statis, artinya hanya menangani data historis dan tidak memperhitungkan dinamika pengguna yang menambahkan ulasan secara real-time. Setelah Netflix mengatasi tantangan ini, mereka menjadikan algoritma pemenang sebagai bagian dari sistem rekomendasinya.

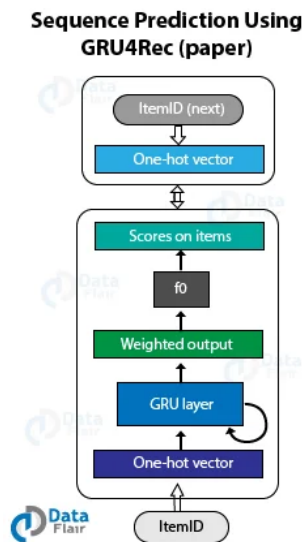
Netflix menggunakan Algoritma Peringkat untuk memberikan daftar peringkat film dan Acara TV yang paling menarik bagi penggunanya. Tetapi, seringkali sulit untuk mengakomodasi semuanya dan menguji kinerjanya secara bersamaan. Untuk mempercepat proses eksperimen algoritma peringkatnya, Netflix menerapkan teknik *interleaving* yang memungkinkannya mengidentifikasi algoritma terbaik. Teknik ini diterapkan dalam dua tahap untuk memberikan algoritma peringkat halaman terbaik demi memberikan rekomendasi yang dipersonalisasi kepada penggunanya. Pada tahap pertama dilakukan eksperimen untuk menentukan *member preference* antara kedua algoritma pemeringkatan tersebut. Berbeda dengan pengujian A/B di mana dua kelompok penonton dihadapkan pada dua algoritma peringkat, Netflix menggunakan *interleaving* untuk memadukan peringkat algoritma A dan B. Netflix menyediakan konten yang mendorong penggunanya berdasarkan teknik *interleaving* yang sangat sensitif terhadap peringkat kualitas algoritma.

Kesadaran Kontekstual adalah salah satu elemen kunci dalam mempersonalisasi rekomendasi untuk penggunanya. Ini tidak hanya meningkatkan kinerja sistem rekomendasi tetapi juga mendorong pengguna untuk memberikan *feedback* yang lebih baik yang akan menghasilkan rekomendasi yang berkualitas. Ada dua kategori kelas kontekstual:

1. Eksplisit, yang terdiri dari:
  - Lokasi
  - Bahasa
  - Waktu dalam hari
  - Perangkat
2. Simpulan, yang terdiri dari:
  - *Binging patterns*
  - Pendamping

Untuk memprediksi konteks, digunakan *representation learning*. Ini merupakan *deep learning technique* yang melakukan rekayasa fitur dan menemukan fitur tanpa pemrograman eksplisit. Berdasarkan waktu dan periode menonton, Netflix mengolah datanya pada berbagai parameter seperti Hari, Minggu, Musim, dan bahkan periode yang lebih lama seperti Olimpiade, FIFA, dan pemilu. Untuk melakukan prediksi kontekstual, Netflix

memperlakukan rekomendasi sebagai masalah klasifikasi urutan. Dibutuhkan input sebagai urutan tindakan pengguna dan melakukan prediksi yang menghasilkan rangkaian tindakan selanjutnya. Contoh masalah urutan adalah Gru4Rec dan dalam kasus prediksi urutan kontekstual, input terdiri dari tindakan pengguna kontekstual serta konteks pengguna saat ini. Ini membantu mesin rekomendasi memecahkan pertanyaan berikut: “Berdasarkan semua tindakan historis yang diambil oleh pengguna, video apa yang paling mungkin mereka putar sekarang?”. Berikut adalah cara Netflix menggunakan Data Science untuk memberikan rekomendasi.



Sumber: DataFlair

Referensi:

Jimy. (2019). *Pengenalan Data Science*. medium.com.

DataFlair Team. (2019). *Data Science at Netflix – A Must Read Case Study for Aspiring Data Scientists*. data-flair.training.

Github Repository: <https://github.com/gekbbb/Studi-Independen-Zenius>