

Prinsip - Prinsip Sistem Informasi

Studi Kasus 5

Kelompok C20:

1. Yosef Nuraga Wicaksana 2206082751
2. Clement Samuel Marly 2206082114
3. Alexander Audric Johansyah 2206815466
4. Tohodo Betrand Simamora 2206083376

Penggunaan Analitik pada Netflix

Para pengguna Netflix menghasilkan banyak informasi detail mengenai minat, selera, dan kebiasaan menonton mereka. Netflix menggunakan data dan analitik untuk menghasilkan rekomendasi tontonan yang disukai pengguna. Netflix juga menggunakan data dan analitik untuk memprediksi permintaan pengguna untuk banyak produksi yang sedang dipertimbangkan untuk ditambahkan. Ketika Netflix membuat kesepakatan dengan para pemeran berbakat seperti Martin Scorsese, Ryan Reynolds, atau Obamas, hal tersebut didasarkan pada model data yang memprediksi peluang keberhasilan kombinasi bakat tertentu, kemungkinan alur cerita, dan faktor lainnya. Netflix tidak hanya dapat memprediksi peluang keberhasilan, tetapi juga dapat memprediksi dengan tingkat akurasi yang tinggi mengenai pelanggan mana yang akan menonton serial tertentu.

Pertanyaan dan Jawaban

1. Tools dan teknologi apa yang mungkin digunakan Netflix untuk menyimpan dan memproses data?

Netflix merupakan salah satu perusahaan yang mengandalkan data penggunanya dalam membentuk sebuah keputusan dimana memerlukan data pengguna yang berskala besar. Oleh karena itu, netflix memerlukan beberapa *tools* yang digunakan dalam mengolah data untuk memberi analisis data yang akurat. Dalam menyimpan data dengan skala besar, netflix memanfaatkan tiga teknologi yang memiliki peran berbeda yakni:

1. Cassandra

Cassandra merupakan teknologi yang berbasis noSQL (*Not only SQL*) yang dapat menyimpan data skala besar dengan performa yang tinggi dengan cepat. Teknologi ini dimanfaatkan oleh netflix dalam menyimpan data dengan tingkat interaksi dengan pengguna yang tinggi secara *realtime* seperti preferensi pengguna, riwayat *playback*, dan data lainnya. Mekanisme penggunaan Cassandra yakni dengan membuat model struktur data terlebih dahulu untuk penyimpanan dan pengambilan data kemudian dioptimalkan dengan mengkonsistenkan distribusi data sehingga tidak ada *bottleneck* yang menimbulkan kesalahan data.

2. Hadoop

Hadoop merupakan teknologi *distributed file system* (penyimpanan dengan jumlah server atau perangkat lebih dari satu yang digunakan oleh netflix dalam menyimpan data yang interaksinya tidak terlalu tinggi atau cepat seperti data performa, data pembayaran, dan data bisnis lainnya. Dalam menggunakan Hadoop netflix akan menyortir data besar kemudian memecah file data tersebut menjadi beberapa data kecil yang didistribusikan ke berbagai cluster dalam infrastruktur Hadoop netflix dengan performa yang tinggi. Selain kemampuannya dalam mendistribusikan data dengan cepat, Hadoop juga memiliki keuntungan

dalam kemampuannya untuk mengintegrasikan beberapa teknologi pemrosesan data seperti Apache Spark dan Apache Hive sehingga efektif dari segi biaya yang diperlukan dalam manajemen data perusahaan.

3. Amazon Simple Storage Service

Amazon Simple Storage Service merupakan salah satu layanan *cloud storage* yang ditawarkan oleh Amazon Web Services. Teknologi ini dimanfaatkan oleh netflix karena skalabilitasnya yang besar dalam menyimpan file netflix yang berukuran besar dan durabilitasnya sehingga data yang tersimpan dalam Amazon S3 dapat diakses kapan saja karena minimnya kemungkinan gangguan. Layanan ini sendiri digunakan dalam menyimpan data dari konten - konten netflix berupa video dan audio hingga profil pengguna netflix.

Data - data yang telah disimpan oleh netflix masih perlu diolah terlebih dahulu sebelum menghasilkan analisis yang berguna bagi perusahaan. Mengingat terdapat banyak data yang disimpan oleh netflix maka diperlukanlah beberapa perangkat dalam pemrosesan data, yakni:

1. Druid

Druid merupakan sebuah perangkat yang digunakan oleh netflix dalam *data query* yang mampu menangani data dengan dimensi tinggi secara cepat sehingga dimanfaatkan dalam eksplorasi dan visualisasi data yang dimiliki oleh perusahaan untuk menjadi bahan pertimbangan keputusan perusahaan.

2. Apache Spark

Apache spark merupakan sebuah perangkat yang digunakan oleh netflix dalam membuat model *machine learning* dari data yang disimpan di Cassandra dan

Hadoop yang mendekati *realtime* dan fleksibel karena memiliki banyak kemampuan dalam pemrosesan berbagai tipe maupun format data sehingga dimanfaatkan dalam proses ETL (*Extract, Transform, Load*) atau proses transformasi data perusahaan.

3. Apache Flink

Apache Flink merupakan perangkat pemrosesan data yang dapat memberikan hasil analisisnya secara *realtime* dengan memerlukan latency yang low sehingga dimanfaatkan oleh netflix dalam prediksi dan deteksi anomali dalam keamanan sistem hingga pemberian rekomendasi konten kepada pengguna karena kecepatannya dalam memproses data dengan sangat cepat.

2. Mengapa Netflix perlu melakukan proses ETL pada semua data?

ETL (*Extract, Transform, Load*) adalah proses mengekstrak data dari berbagai sumber, melakukan penyuntingan pada data tersebut, dan melakukan transformasi dan memuat data yang telah diubah ke dalam *warehouse* data. Proses ETL sangat penting untuk dilakukan bagi Netflix karena mereka mengumpulkan dan memproses jumlah data yang sangat besar dari berbagai sumber. ETL juga membersihkan dan mengorganisir data dengan cara yang mengatasi kebutuhan intelijen bisnis tertentu, seperti pelaporan data bulanan. Selain itu, ETL juga dapat meningkatkan proses *back-end* atau *end user experience*. Berikut adalah tahapan-tahapan dalam proses ETL:

1. *Extract*

Tahap pertama adalah mengumpulkan data dari berbagai sumber, seperti data dari sistem internal Netflix, data dari pihak ketiga, dan data dari media sosial. Data

kemudian diekstrak dari sumbernya, dikonversikan menjadi format tunggal yang cocok untuk diproses, dan disimpan dalam sistem penyimpanan sementara. Data yang tidak memenuhi pola atau nilai yang diharapkan dapat ditolak dari pengolahan selanjutnya.

2. *Transform*

Tahap kedua adalah transformasi data, di mana data yang diekstrak diberikan serangkaian algoritma untuk diubah dan diproses sehingga menjadi lebih mudah diakses dan dipahami. Transformasi melibatkan mengubah format data, membersihkan dan menghapus data yang tidak relevan, membuat atribut baru untuk data, dan menggabungkan data dari beberapa sumber untuk mengurangi waktu pemrosesan yang diperlukan untuk membuat laporan yang diharapkan

3. *Load*

Tahap ketiga adalah memuat data yang telah diubah ke dalam *warehouse* data Netflix. Setelah data dimuat, data dapat diakses dan digunakan untuk membangun model analitik dan untuk memprediksi perilaku pelanggan serta data diperiksa terhadap batasan yang ditentukan dalam skema database untuk memastikan kualitasnya. Tahap pengisian data untuk gudang data yang besar dapat memakan waktu beberapa hari.

Secara keseluruhan, ETL sangat penting bagi Netflix karena memungkinkan mereka untuk mengekstrak, menyunting, dan memuat data dengan cepat dan efisien dari berbagai sumber, sehingga dapat digunakan untuk menghasilkan rekomendasi yang lebih baik dan prediksi perilaku pelanggan.

3. Netflix baru-baru ini menghilangkan sistem review dengan lima bintang dan berhenti menerima review yang dikirimkan pengguna. Sebagai gantinya, sekarang Netflix menggunakan fitur "thumbs up/thumbs down" yang lebih sederhana. Netflix juga menerapkan skor yang dipersonalisasi dengan tujuan untuk mencocokkan penonton dengan konten ideal berdasarkan kebiasaan menonton mereka. Perubahan ini hampir berlawanan dengan intuisi organisasi yang ingin memahami minat penonton. Berikan penjelasan Anda terkait perubahan ini!

Berdasarkan artikel dari Nathan McAlone (2017) yang memberikan pertanyaan kepada *Director of Product Netflix*, Cameron Johnson, sistem penilaian bintang di Netflix merupakan sistem yang dipersonalisasi. Arti dari sistem yang dipersonalisasi adalah ketika ada film dengan rating 4 di Netflix, hal tersebut bukan berarti rata - rata dari semua penilaian orang adalah 4 bintang. Hal tersebut berarti Netflix memperkirakan bahwa seseorang akan memberikan rating 4 pada film tersebut berdasarkan kebiasaan orang tersebut dan orang lainnya yang mirip. Perbedaan kegunaan sistem rating Netflix dengan platform - platform e-commerce lainnya membuat banyak orang salah mengira kegunaan dari sistem rating tersebut (McAlone, 2017).

Kesalahpahaman terkait sistem rating Netflix ditambah dengan banyaknya orang yang tidak termotivasi dalam menggunakan sistem rating Netflix membuat Netflix mengubah sistem rating mereka. Berdasarkan artikel dari Nathan McAlone, Netflix ingin memastikan bahwa sistem rating yang baru dapat dipahami oleh pengguna dengan jelas. Maka dari itu, Netflix memilih sistem "*thumbs up/thumbs down*," yang dapat dipahami secara mudah oleh pengguna dan dapat digunakan untuk melatih algoritma agar mengetahui apa yang pengguna sukai. "Perubahan sederhana ini menyebabkan

peningkatan lebih dari 200% dalam jumlah penilaian dari pengguna," kata Cameron Johnson, *Director of Product Netflix*. Dengan perubahan ini, pengguna juga memiliki cara untuk mencegah Netflix mensugesti sebuah film dengan cara memberi "*thumbs down*" pada suatu film (McAlone, 2017).

Sebagai pengganti dari sistem rating awal Netflix, Netflix juga menambahkan cara untuk mempersonalisasi sugesti film kepada setiap pengguna. Netflix menggunakan algoritma yang menghitung skor setiap film terhadap kebiasaan menonton pengguna. Apabila film tersebut sesuai dengan kebiasaan menonton film pengguna, skor tersebut akan semakin tinggi dan sebaliknya untuk film yang kurang sesuai. Penambahan fitur skor dan perubahan sistem rating Netflix bukan berarti Netflix berhenti ingin memahami minat pengguna, melainkan ingin menghilangkan kesalahpahaman terkait sistem rating mereka dan meningkatkan interaksi pengguna terhadap sistem rating mereka sehingga sugesti film dapat lebih dipersonalisasi atau disesuaikan terhadap setiap pengguna.

4. Apakah akan bermanfaat untuk menambah sumber data Netflix dari jaringan media sosial atau pencarian Internet? Nilai tambahan apa yang mungkin didapatkan dari data tersebut?

Menambah sumber data dari jaringan media sosial maupun dari pencarian internet sangatlah bermanfaat bagi Netflix. Berikut manfaat yang dapat didapatkan oleh Netflix jika menambahkan sumber data dari jaringan sosial maupun pencarian internet.

1. Data yang diambil dari media sosial atau dari pencarian internet akan membantu Netflix dalam memprediksi tren dan pola perilaku pengguna Netflix. Data tersebut dapat dimanfaatkan untuk memperkirakan tren yang akan berlaku di

masa yang akan datang dan data tersebut dapat menjadi salah satu pertimbangan untuk produksi baru.

2. Data yang diambil dari media sosial dan pencarian internet dapat membantu Netflix untuk menemukan informasi yang lebih kaya tentang minat pengguna Netflix. Dengan memperoleh informasi yang banyak tentang minat pengguna Netflix, Netflix dapat meningkatkan rekomendasi mereka secara signifikan.
3. Terakhir, data yang diambil dari media sosial atau dari pencarian internet akan membantu Netflix untuk memperluas jangkauan promosinya. Dengan data yang didapat, Netflix dapat memahami bagaimana promosi juga dapat mempengaruhi perilaku pengguna yang ada di luar platform Netflix.

Referensi

Anon. (n.d.). *ETL (Extract, Transform, Load)*. IBM. Diakses pada 12 Maret, 2023, melalui <https://www.ibm.com/id-en/topics/etl>

Harahap, C. (2023, Maret 13). *Business Intelligence: Big Data and Analytics*. Scele.cs.ui.ac.id. Diakses pada 12 Maret, 2023, melalui https://scele.cs.ui.ac.id/pluginfile.php/177602/mod_resource/content/3/Ch06-Business%20Intelligence_Big%20Data%20and%20Analytics%20V1.pdf

Izrailevsky Y.(2016,12 Februari)*Completing the Netflix Cloud Migration - About Netflix*.Netflix. Diakses pada 12 Maret 2023, melalui <https://about.netflix.com/en/news/completing-the-netflix-cloud-migration>

Krishnan S. (2013, Januari 10). *Hadoop Platform as a Service in the Cloud - Netflix TechBlog*. Medium.Diakses pada 13 Maret 2023, melalui <https://netflixtechblog.com/hadoop-platform-as-a-service-in-the-cloud-c23f35f965e7>

McAlone, N. (2017, April 5). *The exec who replaced Netflix's 5-star rating system with 'thumbs up, Thumbs Down' explains why*. Business Insider. Diakses pada 13 Maret, 2023, melalui <https://www.businessinsider.com/why-netflix-replaced-its-5-star-rating-system-2017-4>