**Anonimisasi data**

**Apa itu anonimisasi data?**

Anda telah belajar tentang pentingnya privasi dalam analisis data. Sekarang, saatnya membahas tentang **anonimisasi data** dan jenis data apa yang harus dianonimkan. **Informasi pengenal pribadi** , atau **PII** , adalah informasi yang dapat digunakan sendiri atau dengan data lain untuk melacak identitas seseorang.

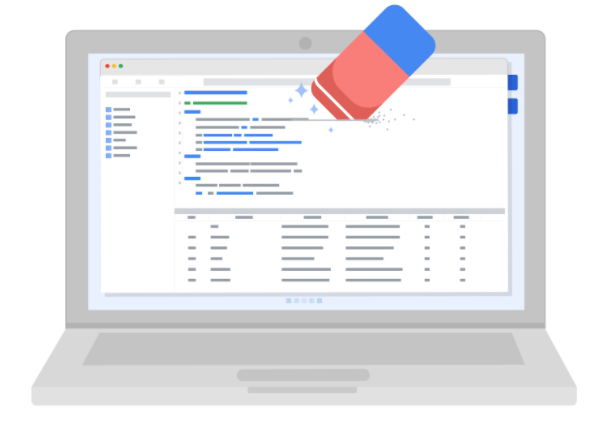
Anonimisasi data adalah proses melindungi data pribadi atau sensitif seseorang dengan menghilangkan informasi semacam itu. Biasanya, anonimisasi data melibatkan pengosongan, hashing, atau penyembunyian informasi pribadi, sering kali dengan menggunakan kode dengan panjang tetap untuk mewakili kolom data, atau menyembunyikan data dengan nilai yang diubah.

**Peran Anda dalam anonimisasi data**

Organisasi memiliki tanggung jawab untuk melindungi data mereka dan informasi pribadi yang mungkin terkandung dalam data tersebut. Sebagai seorang analis data, Anda mungkin diharapkan memahami data apa saja yang perlu dianonimkan, namun secara umum Anda tidak bertanggung jawab atas anonimisasi data itu sendiri. Pengecualian yang jarang terjadi mungkin terjadi jika Anda bekerja dengan salinan data untuk tujuan pengujian atau pengembangan. Dalam hal ini, Anda mungkin diminta untuk menganonimkan data sebelum Anda mengerjakannya.

**Jenis data apa yang harus dianonimkan?**

Data layanan kesehatan dan keuangan adalah dua jenis data yang paling sensitif. Industri-industri ini sangat bergantung pada teknik anonimisasi data. Lagipula, taruhannya sangat tinggi. Itu sebabnya data di kedua industri ini biasanya melalui **de-identifikasi** , yaitu **proses yang digunakan untuk menghapus data dari semua informasi identitas pribadi** .



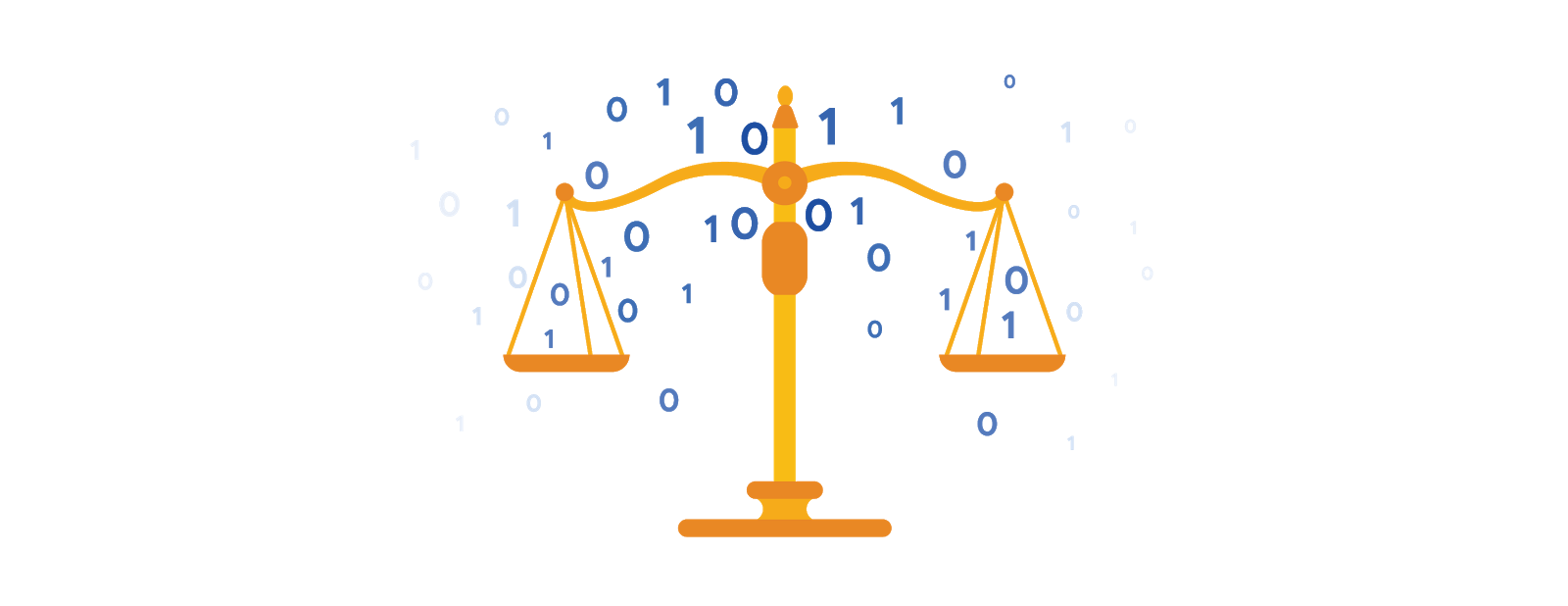
Anonimisasi data digunakan di hampir setiap industri. Itulah mengapa sangat penting bagi analis data untuk memahami dasar-dasarnya. Berikut daftar data yang sering dianonimkan:

* Nomor telepon
* Nama
* Plat nomor dan nomor lisensi
* Nomor jaminan sosial
* alamat IP
* Rekam medis
* Alamat email
* Foto-foto
* Nomor rekening

Bagi sebagian orang, masuk akal jika jenis data ini harus dianonimkan. Bagi yang lain, kami harus sangat spesifik tentang apa yang perlu dianonimkan. Bayangkan sebuah dunia di mana kita semua memiliki akses ke alamat, nomor rekening, dan informasi identitas lainnya satu sama lain. Hal ini akan mengganggu privasi banyak orang dan membuat dunia menjadi kurang aman. Anonimisasi data adalah salah satu cara kami menjaga privasi dan keamanan data!

**Perdebatan data terbuka**

Sama seperti privasi data, data terbuka adalah topik yang banyak diperdebatkan di dunia saat ini. Analis data banyak memikirkan data terbuka, dan sebagai analis data masa depan, Anda perlu memahami dasar-dasar agar berhasil dalam peran baru Anda.



**Apa itu data terbuka?**

Dalam analisis data, **data terbuka** adalah bagian dari **etika data,** yang berkaitan dengan penggunaan data secara etis. **Keterbukaan** mengacu pada akses bebas, penggunaan, dan berbagi data. Namun agar data dianggap terbuka, data tersebut harus:

* Tersedia dan dapat diakses oleh publik sebagai kumpulan data yang lengkap
* Disediakan berdasarkan ketentuan yang memungkinkan untuk digunakan kembali dan didistribusikan kembali
* Izinkan partisipasi universal sehingga siapa pun dapat menggunakan, menggunakan kembali, dan mendistribusikan ulang data

Data hanya dapat dianggap terbuka jika memenuhi ketiga standar tersebut.

**Perdebatan data terbuka: Data apa yang harus tersedia untuk umum?**

Salah satu manfaat terbesar dari data terbuka adalah database yang kredibel dapat digunakan secara lebih luas. Pada dasarnya, ini berarti semua data bagus tersebut dapat dimanfaatkan, dibagikan, dan digabungkan dengan data lain. Hal ini dapat berdampak besar pada kolaborasi ilmiah, kemajuan penelitian, kapasitas analitis, dan pengambilan keputusan. Namun penting untuk memikirkan individu-individu yang diwakili oleh publik, juga data terbuka.

**Data pihak ketiga** dikumpulkan oleh entitas yang tidak memiliki hubungan langsung dengan data tersebut. Anda mungkin ingat pernah mempelajari jenis data ini sebelumnya. Misalnya, pihak ketiga mungkin mengumpulkan informasi tentang pengunjung situs web tertentu. Dengan melakukan hal ini, pihak ketiga dapat membuat profil audiens, yang membantu mereka lebih memahami perilaku pengguna dan menargetkan mereka dengan iklan yang lebih efektif.

**Informasi pengidentifikasi pribadi (PII)** adalah data yang kemungkinan besar dapat mengidentifikasi seseorang dan membuat informasi tentang orang tersebut diketahui. Penting untuk menjaga keamanan data ini ***.*** PII dapat mencakup alamat seseorang, informasi kartu kredit, nomor jaminan sosial, catatan medis, dan banyak lagi.

Semua orang ingin merahasiakan informasi pribadi tentang dirinya. Karena data pihak ketiga sudah tersedia, penting untuk menyeimbangkan keterbukaan data dengan privasi individu.

# Situs dan sumber daya untuk data terbuka

Beruntung bagi analis data, ada banyak situs dan sumber terpercaya yang tersedia untuk data terbuka. Penting untuk diingat bahwa bahkan data yang bereputasi baik pun perlu terus dievaluasi, namun situs web berikut adalah titik awal yang berguna:

1. [**Situs data pemerintah AS**](https://www.data.gov/): Data.gov adalah salah satu sumber data terlengkap di AS. Sumber daya ini memberi pengguna data dan alat yang mereka perlukan untuk melakukan penelitian, dan bahkan membantu mereka mengembangkan aplikasi web dan seluler serta merancang visualisasi data.
2. [**Biro Sensus AS**](https://www.census.gov/data.html): Sumber data terbuka ini menawarkan informasi demografis dari pemerintah federal, negara bagian, dan lokal, serta entitas komersial di AS.
3. [**Buka Jaringan Data**](https://www.opendatanetwork.com/): Sumber data ini memiliki mesin pencari yang sangat canggih dan filter tingkat lanjut. Di sini, Anda dapat menemukan data tentang topik-topik seperti keuangan, keselamatan publik, infrastruktur, serta perumahan dan pembangunan.
4. [**Kumpulan Data Publik Google Cloud**](https://cloud.google.com/public-datasets): Ada sejumlah kumpulan data publik pilihan yang tersedia melalui Program Kumpulan Data Publik Google Cloud yang Anda temukan sudah dimuat ke BigQuery.
5. [**Pencarian Kumpulan Data**](https://datasetsearch.research.google.com/)**:** Pencarian Kumpulan Data adalah mesin pencari yang dirancang khusus untuk kumpulan data; Anda dapat menggunakan ini untuk mencari kumpulan data tertentu.