**Jelajahi privasi diferensial**

Ilmuwan data memiliki tanggung jawab etis (dan seringkali legal) untuk melindungi data sensitif. Privasi diferensial adalah pendekatan terdepan yang memungkinkan analisis berguna sekaligus melindungi nilai data yang dapat diidentifikasi secara individual.

**Tujuan pembelajaran**

Setelah menyelesaikan modul ini, Anda dapat:

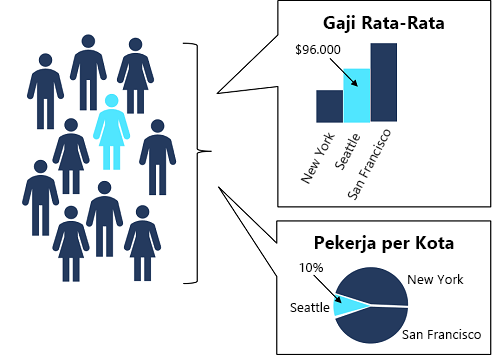
* Mengartikulasikan masalah privasi data
* Jelaskan cara kerja privasi diferensial
* Konfigurasikan parameter untuk privasi diferensial
* Lakukan analisis data pribadi yang berbeda

1. **Pengantar**

Proyek ilmu data, termasuk proyek pembelajaran mesin, melibatkan analisis data; dan seringkali data tersebut menyertakan detail pribadi yang sensitif yang harus dijaga kerahasiaannya. Dalam praktiknya, sebagian besar laporan yang dipublikasikan dari data menyertakan Himpunan Data, yang menurut Anda akan memberikan privasi – lagipula, hasil gabungan tidak mengungkapkan nilai data individual.

Namun, pertimbangkan kasus di mana beberapa analisis data menghasilkan agregasi yang dilaporkan yang bila digabungkan, dapat digunakan untuk mencari informasi tentang individu dalam himpunan data sumber. Misalkan 10 peserta berbagi data tentang lokasi dan gaji mereka, dari mana dua laporan dihasilkan:

* Laporan gaji gabungan yang memberi tahu kami gaji rata-rata di New York, San Francisco, dan Seattle
* Laporan lokasi pekerja yang memberi tahu kita bahwa 10% dari peserta studi (dengan kata lain, satu orang) tinggal di Seattle.



Dari dua laporan ini, kita dapat dengan mudah menentukan gaji spesifik dari peserta yang berbasis di Seattle. Siapa pun yang meninjau kedua studi yang kebetulan mengenal seseorang dari Seattle yang berpartisipasi, sekarang tahu gaji orang itu.

Dalam modul ini, Anda akan menjelajahi **privasi diferensial**, sebuah teknik yang dapat membantu melindungi data individu dari paparan semacam ini.

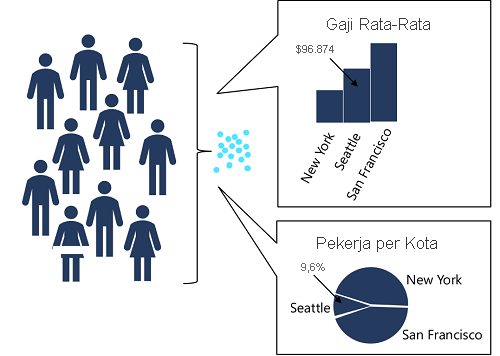
**Tujuan pembelajaran**

Dalam modul ini, Anda akan mempelajari cara:

* Mengartikulasikan masalah privasi data
* Jelaskan cara kerja privasi diferensial
* Konfigurasikan parameter untuk privasi diferensial
* Lakukan analisis data pribadi yang berbeda

1. **Memahami privasi diferensial**

Privasi diferensial berupaya melindungi nilai data individual dengan menambahkan "gangguan" statistik ke dalam proses analisis. Matematika yang terlibat dalam menambahkan kebisingan itu rumit, tetapi prinsipnya cukup intuitif - kebisingan memastikan bahwa agregasi data tetap konsisten secara statistik dengan nilai data aktual yang memungkinkan untuk beberapa variasi acak, tetapi membuatnya tidak mungkin untuk menghitung nilai individu dari agregat data. Selain itu, noise berbeda untuk setiap analisis, sehingga hasilnya non-deterministik – dengan kata lain, dua analisis yang melakukan agregasi yang sama dapat menghasilkan hasil yang sedikit berbeda.



1. **Konfigurasikan parameter privasi data**

Salah satu cara agar seseorang dapat melindungi data pribadinya adalah dengan tidak berpartisipasi dalam penelitian – ini dikenal sebagai opsi "memilih tidak ikut". Namun, ada beberapa pertimbangan untuk ini sebagai solusi:

* Bahkan jika Anda memutuskan untuk tidak ikut penelitian, penelitian mungkin masih menghasilkan hasil yang memengaruhi Anda. Misalnya, Anda dapat memilih untuk tidak mengikuti studi yang membandingkan diagnosis penyakit jantung pada sekelompok orang atas dasar bahwa hal itu dapat mengungkapkan diagnosis penyakit jantung yang menyebabkan premi asuransi kesehatan Anda meningkat. Jika penelitian menemukan korelasi antara orang yang minum kopi dan risiko penyakit jantung yang lebih tinggi, dan perusahaan asuransi Anda mengetahui bahwa Anda adalah seorang peminum kopi, tingkat Anda mungkin naik meskipun Anda tidak berpartisipasi secara pribadi dalam penelitian ini.
* Manfaat partisipasi dalam penelitian ini mungkin lebih besar daripada dampak negatifnya. Misalnya, jika Anda dibayar $100 untuk berpartisipasi dalam studi yang mengakibatkan tingkat asuransi kesehatan Anda naik $10 per tahun, itu akan lebih dari 10 tahun sebelum Anda mengalami kerugian bersih. Ini mungkin merupakan tradeoff yang berharga bagi Anda (terutama jika tarif Anda mungkin naik sebagai hasil dari penelitian ini bahkan jika Anda tidak berpartisipasi!)
* Satu-satunya cara agar opsi opt-out bekerja untuk setiap individu, adalah agar setiap individu tidak ambil bagian – yang membuat keseluruhan studi menjadi sia-sia!

Jumlah variasi yang disebabkan oleh penambahan noise dapat dikonfigurasi melalui parameter yang disebut epsilon. Nilai ini mengatur jumlah risiko tambahan bahwa data pribadi Anda dapat diidentifikasi dengan menolak opsi keluar dan berpartisipasi dalam penelitian. Kuncinya adalah menerapkan prinsip privasi ini untuk semua orang yang berpartisipasi dalam penelitian ini. Nilai epsilon yang rendah memberikan privasi paling tinggi, dengan mengorbankan akurasi yang lebih rendah saat menggabungkan data. Nilai epsilon yang lebih tinggi menghasilkan agregasi yang lebih sesuai dengan distribusi data aktual, tetapi di mana kontribusi individu dari satu individu ke nilai agregat kurang dikaburkan oleh noise.

IMG_256

1. **Latihan - Gunakan privasi diferensial**

Sekarang adalah kesempatan Anda untuk menjelajahi privasi diferensial untuk diri Anda sendiri dengan menggunakan paket **SmartNoise**.

Dalam latihan ini, Anda akan:

* Gunakan **SmartNoise** untuk menghasilkan analisis pribadi yang berbeda.
* Gunakan **SmartNoise** untuk mengirimkan kueri pribadi yang berbeda.

1. **Instruksi**

Ikuti petunjuk ini untuk menyelesaikan latihan.

1. Jika Anda belum memiliki langganan Azure, daftar untuk uji coba gratis di <https://azure.microsoft.com.>
2. Menampilkan repo latihan di <https://aka.ms/mslearn-dp100.> \ <https://microsoftlearning.github.io/mslearn-dp100/>
3. Jika Anda belum melakukannya, selesaikan latihan **Membuat ruang kerja Azure Machine Learning** untuk memprovisikan ruang kerja Azure Machine Learning, membuat instans komputasi, dan mengkloning file yang diperlukan.
4. Selesaikan latihan **Jelajahi privasi diferensial.**

**5. Ringkasan**

Dalam modul ini, Anda belajar cara:

* Mengartikulasikan masalah privasi data
* Jelaskan cara kerja privasi diferensial
* Konfigurasikan parameter untuk privasi diferensial
* Lakukan analisis data pribadi yang berbeda

Untuk mempelajari lebih lanjut tentang menafsirkan model, lihat **Privasi Diferensial** dalam dokumentasi Azure Machine Learning.

**Link** : <https://learn.microsoft.com/id-id/azure/machine-learning/concept-responsible-ai#privacy-and-security>