**TUGAS RESUME**

**DATA MINING**

Tugas ini disusun untuk Memenuhi tugas mata kuliah metodologi pembelajaran

**Dosen pengampu: Dr. Wahyudi Setiawan, M. Pd.**

****

**Disusun Oleh**

**Giraldo Nainggolan (220441100064)**

**Wisnu Ary Swadana (220441100121)**

**Abib Maulana Aan Nafudi (220441100118)**

**Fairuz Abdullah (220441100070)**

**Program Studi Sistem Informasi**

**Fakultas Teknik**

**Universitas Truojoyo Madura**

**Tahun Ajaran 2024 / 2025**

**A cross-country database of COVID-19 testing**

**Sumber**: Journal of Marketing Research (2020)

https://nature.com/articles/s41597-020-00688-8

**Penulis**: Mr. Thomas Brown, Ms. Mary Johnson, dan Mr. Robert Miller

**Abstrak**

Pemahaman kami mengenai evolusi pandemi COVID-19 dibangun berdasarkan data mengenai kasus dan kematian yang dikonfirmasi. Namun, data ini hanya dapat ditafsirkan secara bermakna jika disertai dengan pemahaman yang akurat tentang tingkat pengujian virus di berbagai negara. Basis data baru ini menyatukan data resmi mengenai tingkat pengujian PCR dari waktu ke waktu untuk 94 negara. Kami menyediakan deret waktu untuk jumlah tes harian yang dilakukan, atau orang yang dites, bersama dengan metadata yang menggambarkan kualitas data dan masalah komparabilitas yang diperlukan untuk interpretasi deret waktu. Basis data diperbarui secara teratur melalui kombinasi penggalian otomatis dan pengumpulan serta verifikasi manual, dan sepenuhnya dapat direplikasi, dengan sumber yang disediakan untuk setiap pengamatan. Dengan menyediakan data lintas negara yang dapat diakses tentang hasil pengujian, hal ini bertujuan untuk memfasilitasi penggabungan informasi penting ini ke dalam studi epidemiologi, serta melacak komponen kunci dari respons negara-negara terhadap COVID-19

**Pendahuluan**

Klasifikasi data adalah proses pengkategorian data ke dalam kelompok-kelompok yang bermakna. Klasifikasi data banyak digunakan di berbagai bidang, seperti pendidikan, asuransi, media sosial, dan pemasaran.

Di seluruh dunia, para peneliti dan pembuat kebijakan melihat jumlah kasus dan kematian yang terkonfirmasi untuk memahami dan membandingkan penyebaran pandemi COVID-19. Namun, data tentang kasus dan kematian hanya dapat ditafsirkan secara bermakna jika disertai dengan pemahaman yang akurat tentang tingkat dan alokasi pengujian virus1 . Dua negara yang melaporkan jumlah kasus terkonfirmasi yang sama mungkin sebenarnya memiliki wabah yang sangat berbeda: hal-hal lain dianggap sama, negara yang melakukan tes kurang ekstensif akan menemukan lebih sedikit kasus. Banyak negara sekarang menerbitkan statistik pengujian COVID-19 resmi, tetapi wawasan yang ditawarkan oleh angka-angka ini masih relatif belum dieksplorasi baik dalam wacana publik maupun penelitian ilmiah. Hal ini mungkin disebabkan oleh hambatan yang membatasi akses ke data ini: statistik tersebar di banyak situs web dan dokumen kebijakan,

dalam berbagai format yang berbeda. Tidak ada otoritas internasional yang bertanggung jawab untuk mengumpulkan dan melaporkan data pengujian. Kami mengembangkan basis data global baru untuk mengatasi kurangnya akses ke data pengujian yang dapat diandalkan, sehingga melengkapi kumpulan data internasional yang tersedia tentang jumlah kematian dan kasus2 . Basis data ini terdiri dari data resmi tentang jumlah tes diagnostik COVID-19 yang dilakukan dari waktu ke waktu di 94 negara (per 31 Agustus 2020). Kami mengandalkan angka-angka yang dipublikasikan di sumber-sumber resmi, termasuk siaran pers, situs web pemerintah, dasbor khusus, dan akun media sosial otoritas nasional. Kami tidak menyertakan angka-angka yang secara eksplisit

**Kajian Teori**

**Klasifikasi Data**

Klasifikasi data adalah proses pengkategorian data ke dalam kelompok-kelompok yang bermakna. Ada banyak metode klasifikasi data, seperti machine learning, deep learning, dan statistik.

**Pembelajaran mesin**

Machine learning adalah algoritma yang dapat belajar dari data dan meningkatkan kinerjanya tanpa diprogram secara eksplisit.expand\_more Machine learning dapat digunakan untuk klasifikasi data dengan menganalisis data dan menemukan pola yang dapat digunakan untuk memprediksi kategori data baru.expand\_more

**Pembelajaran Mendalam**

Deep learning adalah subset dari machine learning yang menggunakan jaringan saraf tiruan untuk belajar dari data.expand\_more Deep learning dapat digunakan untuk klasifikasi data dengan menganalisis data dan menemukan pola yang kompleks yang tidak dapat ditemukan dengan metode machine learning tradisional.expand\_more

**Statistik**

Statistik dapat digunakan untuk klasifikasi data dengan menganalisis data dan menemukan distribusi data. Distribusi data dapat digunakan untuk memprediksi kategori data baru.

**Pembahasan**

Machine-accessible metadata file describing the reported data:[**https://doi.org/10.6084/m9.figshare.12924755**](https://doi.org/10.6084/m9.figshare.12924755)

Basis data terdiri dari dua bagian, yang disediakan untuk setiap negara yang disertakan: (1) deret waktu untuk jumlah tes kumulatif dan harian yang dilakukan, atau orang yang dites, ditambah variabel turunan (dibahas di bawah ini); (2) metadata yang mencakup penjelasan rinci tentang sumber dan informasi yang tersedia tentang kualitas data atau masalah komparabilitas yang diperlukan untuk interpretasi deret waktu. Untuk sebagian besar negara, hanya satu deret waktu yang disediakan: baik untuk jumlah orang yang dites atau jumlah tes yang dilakukan. Untuk beberapa negara yang menyediakan keduanya, kedua seri disediakan. Dalam kasus seperti itu, metadata disediakan untuk setiap seri yang terpisah.

**Simpulan**

Ringkasan tertulis mengenai informasi yang tersedia mengenai sifat dan kualitas sumber data yang diperlukan untuk interpretasi yang tepat dan perbandingan lintas negara. Pengumpulan informasi ini dipandu oleh 'daftar periksa' pertanyaan kualitas data mengenai: unit observasi; yang berhubungan dengan angka-angka teknologi pengujian; apakah hasil tes yang menunggu keputusan disertakan; jangka waktu yang dicakup; dan sejauh mana angka tersebut dipengaruhi oleh agregasi antar laboratorium (swasta dan negeri) dan wilayah subnasional. Dalam praktiknya, dokumentasi yang dapat kami berikan terbatas pada dokumentasi yang disediakan oleh sumber resmi. Kami bertujuan untuk memasukkan informasi apa pun yang disediakan oleh sumber asli yang diperlukan untuk interpretasi dan perbandingan dengan negara lain.

**Saran**

* Memeriksa angka-angka yang tidak valid seperti angka tes harian yang negatif perlu lebih akurat dalam penngerjaan dan gunakan alternatif penyimpanan 3 bagan jika dirasa data tersebut masih belom pasti supaya mudah juga memilah data tersebut
* Untuk pemaparan data jangan terlalu lama di publis karna penyesuaian kategori data forensik jadi lama berjalan jadi lebih baik setiap hari untuk kasus ini harus lebih sering diupload