

Madani: Jurnal Ilmiah Multidisiplin

Volume 1, Nomor 11, December 2023

Licensed by CC BY-SA 4.0

E-ISSN: [2986-6340](https://doi.org/10.2986-6340)

DOI: <https://doi.org/10.5281/zenodo.10408398>

Modernisasi *Official Statistic* Dengan *Big Data*

Hotton Jonatan¹

¹Program Studi DIII Statistik, Politeknik Statistik STIS,

Jalan Otto Iskandardinata 64C Jakarta 13330, Indonesia

Email: 112212648@stis.ac.id

Abstrak

Dalam makalah ini, kami menjelaskan dan membahas peluang modernisasi statistik resmi melalui big data. Big data hadir dalam volume tinggi, kecepatan tinggi, dan variasi tinggi. Volume tinggi dapat menghasilkan akurasi yang lebih baik dan lebih banyak detail, kecepatan tinggi dapat menghasilkan estimasi statistik yang lebih sering dan lebih tepat waktu, dan variasi tinggi dapat memberikan peluang untuk statistik di bidang baru. Namun, ada juga banyak tantangan: terdapat perubahan yang tidak terkendali dalam sumber yang mengancam kelangsungan dan perbandingan, dan data yang hanya secara tidak langsung terkait dengan fenomena yang menarik dari segi statistik. Selain itu, big data mungkin sangat fluktuatif dan selektif: cakupan populasi yang mereka referensikan dapat berubah setiap hari, menyebabkan lonjakan yang tidak dapat dijelaskan dalam deret waktu. Dan seringkali, observasi individual dalam set data big data ini tidak memiliki variabel yang memungkinkan mereka dihubungkan ke set data atau kerangka populasi lainnya. Hal ini sangat membatasi kemungkinan koreksi selektivitas dan fluktuasi. Juga, dengan kemajuan big data dan open data, ada lebih banyak peluang untuk pengungkapan data individu, dan ini menimbulkan masalah baru bagi lembaga statistik. Jadi, big data dapat dianggap sebagai sampel nonprobabilitas. Penggunaan sumber-sumber tersebut dalam statistik resmi memerlukan pendekatan lain daripada pendekatan tradisional berdasarkan survei dan sensus. Pendekatan pertama adalah menerima big data apa adanya: indikator yang tidak sempurna, namun sangat tepat waktu, dari perkembangan dalam masyarakat. Pendekatan kedua adalah menggunakan model formal dan mengekstrak informasi dari data ini. Dalam beberapa tahun terakhir, banyak metode baru untuk menangani big data telah dikembangkan oleh ahli statistik matematika dan terapan.

Kata Kunci: Big Data, Official Statistics (Statistik Resmi), NSO (National Statistic Office)/Lembaga Statistik Nasional

Abstract

In this paper, we describe and discuss opportunities for modernizing official statistics through big data. Big data comes in high volume, high speed, and high variety. High volume can lead to better accuracy and more detail, high speed can lead to more frequent and more timely statistical estimates, and high variation can provide opportunities for statistics in new areas. However, there are also many challenges: there are uncontrollable changes in sources that threaten continuity and comparability, and data that are only indirectly related to phenomena of statistical interest. In addition, big data may be highly volatile and selective: the scope of the population they reference can change daily, causing unexplained spikes in the time series. And often, individual observations in these big data datasets do not have variables that allow them to be linked to other datasets or population frames. This severely limits the possibility of selectivity and fluctuation correction. Also, with the advancement of big data and open data, there are more opportunities for individual data disclosure, and this poses new problems for statistical agencies. So, big data can be considered a nonprobability sample. The use of these sources in official statistics requires a different approach than the traditional one based on surveys and censuses. The first approach is to accept big data for what it is: an imperfect, but very timely, indicator of developments in society. The second approach is to use formal models and extract information from this data. In recent years, many new methods for handling big data have been developed by mathematical and applied statisticians. National statistical agencies have always been reluctant to use models, except in special cases such as small area estimation.

Keywords : Big Data, Official Statistics, NSO (National Statistical Office)

Article Info

Received date: 28 November 2023

Revised date: 05 December 2023

Accepted date: 15 December 2023

PENDAHULUAN

Perbincangan mengenai *Big Data* dapat memengaruhi dari *official statistic* atau statistik resmi menjadi salah satu bentuk dari “Modernisasi Data” menjadi lebih baik yang sedang berlangsung. Perbincangan ini menarik perhatian serta memunculkan beberapa kali forum publikasi atau diskusi. Belakangan ini juga artikel-artikel referensi mengenai Revolusi *Official Statistic* dengan *Big Data* muncul dengan pembahasan yang sama yaitu mengenai solusi serta potensial *Big Data* yang akan mempercepat serta membuat revolusi dalam produksi *Official Statistic*.

Big data datang dalam volume besar, kecepatan tinggi, dan variasi tinggi. Contohnya adalah *web scraping*, pesan Twitter, Facebook, catatan ponsel, data loop lalu lintas, dan transaksi perbankan. Ini memberikan peluang untuk statistik baru atau redesain atau modernisasi statistik yang sudah ada. Potensi big data untuk statistik resmi terletak pada jumlah informasi yang luar biasa. Misalnya, ukurannya yang besar mungkin memungkinkan penambahan detail lebih lanjut pada statistik resmi. Volume besar mereka dapat menghasilkan akurasi yang lebih baik dan detail lebih banyak, kecepatan tinggi mereka dapat menghasilkan perkiraan statistik yang lebih sering dan lebih tepat waktu, dan variasi tinggi mereka dapat menciptakan statistik di area baru.

Dari perspektif metodologis, *big data* menimbulkan banyak tantangan. *Big data* dapat sangat volatil dan selektif. Cakupan populasi yang mereka dapat berubah dari hari ke hari, menyebabkan lonjakan yang tidak dapat dijelaskan dalam deret waktu. Seringkali observasi individu dalam set data dalam *big data* kurang memiliki variabel pengait dan oleh karena itu tidak dapat dihubungkan dengan set data atau kerangka populasi lainnya. Ini sangat membatasi kemungkinan koreksi selektivitas dan volatilitas. Berdasarkan penjelasan di atas, secara khusus kita melihat pertanyaan utama yaitu mengenai strategi terbaik apa untuk menggunakan dan modernisasi *big data* dalam statistik resmi?

Tujuan penulisan ini kurang lebih akan membahas yang pertama sumber daya big data dan penggunaan yang mungkin. Kemudian pembahasan mengenai statistik dari big data saat dikumpulkan atau dirangkai, yaitu sebagai statistik yang sah dan menjelaskan beberapa contoh big data sebagai statistik resmi. Kemudian akan membahas model dari modernisasi yang berguna untuk menciptakan informasi dari sumber daya *big data* agar NSO (*National Statistik Office*) atau lembaga statistik nasional dapat menggunakan model untuk menciptakan statistik resmi. Kemudian terakhir adalah pembahasan isu big data saat ini serta posisi big data di masa depan.

METODE

Artikel ini merupakan artikel konseptual yaitu menggunakan metode literatur-literatur yang ada dan hanya menggunakan metode kajian pustaka yang berkenaan dengan metode pengumpulan data pustaka, membaca dan mencatat, serta mengolah bahan penelitian. (Zed, 2008)

HASIL DAN PEMBAHASAN

Sumber daya big data dan penggunaan yang mungkin

Data Lalu-Lintas

Sumber daya big data yang pertama adalah data lalu-lintas kendaraan misalnya data dari Korlantas [Polri](#) yang diolah Pusiknas Bareskrim Polri menunjukkan, jumlah pelanggaran [lalu lintas](#) (lalin) yang terekam melalui Electronic Traffic Law Enforcement (ETLE) yaitu ETLE yang akan menindak pelanggaran secara otomatis melalui kamera pemantau. Dari ELTE ini pelanggaran tercatat mencapai 512,9 ribu kendaraan sepanjang Januari-Mei 2023. “Sampai hari ini, jumlah kamera ETLE adalah 433 jenis statis, lima *weight in motion* atau untuk penimbangan bersifat *mobile* kemudian 806 *mobile handheld* dan 65 *mobile on-board*,” Irjen Pol Firman Shantyabudi, Kepala Korps Lalu Lintas Polri di Jakarta (05/07/2023).

Pesan Media Sosial

Media sosial adalah sumber data di mana orang secara sukarela berbagi informasi, mendiskusikan topik-topik yang menarik, dan berhubungan dengan keluarga dan teman-teman. Berdasarkan data dari *We Are Social* pengguna media sosial Indonesia mencapai 191,4 juta yaitu sekitar 68% dari jumlah penduduk Indonesia. Rinciannya adalah pengguna Instagram yang mencapai 99.15 juta, Youtube 128 juta, Tiktok 92 juta dan Twitter serta aplikasi lainnya 18 Juta pengguna. Jumlah pengguna yang besar menciptakan volume data yang sangat besar serta bervariasi sehingga pesan serta data media sosial menjadi salah satu sumber daya dari big data yang bisa digunakan oleh NSO atau pihak swasta untuk meningkatkan kualitas statistik mereka

Data ponsel

Saat ini, orang-orang membawa ponsel genggam dan menggunakannya sepanjang hari sehingga untuk mengelola lalu lintas telepon, banyak data perlu diproses oleh perusahaan ponsel genggam. Data ini sangat erat kaitannya dengan perilaku orang, yaitu perilaku yang menarik untuk statistik resmi. Sebagai contoh, lalu lintas telepon diarahkan melalui menara telepon yang tersebar geografis, yang memungkinkan penentuan lokasi pengguna telepon. Menara pemancar, bagaimanapun, dapat berubah beberapa kali selama panggilan: algoritma lokasi yang tidak sederhana diperlukan. (*Braaksma & Zeelenberg, 2020*)

Kerangka Kualitas Statistik Big Data

Mengikuti Hukum Statistik UE, yaitu kualitas statistik (relevansi, akurasi, ketepatan waktu, aksesibilitas, perbandingan, dan koherensi) dan prinsip statistik (independensi, ketidakberpihakan, objektivitas, keandalan, kerahasiaan, dan efektivitas biaya. Fokus pembahasan bab ini adalah akurasi, objektivitas, dan keandalan, karena ini merupakan hal yang fundamental untuk statistik resmi. Jika statistik tidak cukup akurat dalam menggambarkan masyarakat atau tidak objektif atau dapat diandalkan, maka statistik tersebut pada dasarnya tidak berguna. Kemudian akan membahas relevansi, ketepatan waktu, aksesibilitas, perbandingan, dan koherensi.

Akurasi

Akurasi statistik dilihat dari varians dan biasnya. Untuk survei berdasarkan sampel probabilitas, biasanya sekitar nol, tetapi variansnya positif, sedangkan untuk sensus varians dan biasnya sekitar nol. *Big data* mungkin sangat fluktuatif, selektif dan cakupan populasi yang mereka referensikan dapat berubah dari hari ke hari, menyebabkan lonjakan tidak dapat dijelaskan dalam deret waktu tertentu. Oleh karena itu, perlu menggunakan informasi tambahan, baik dari sumber *big data* itu sendiri atau dari sumber data lain. Informasi tambahan tersebut harus cukup untuk koreksi bias sehingga statistik tersebut akan relevan bagi pengguna. Informasi ini dapat terdiri dari karakteristik individu dalam kumpulan data yang membuat bagian penting dari populasi yang penting hilang sepenuhnya

Objektivitas dan Tahan Uji

Prinsip objektivitas berarti bahwa data yang digunakan untuk mengestimasi model harus merujuk pada fenomena yang sedang dijelaskan; dengan kata lain, objek dan populasi untuk model tersebut sesuai dengan fenomena statistik yang sedang berlangsung. Prinsip tahan uji berarti harus mencegah revisi data statistik resmi hanya karena model berubah, misalnya, karena kegagalan model. Terutama untuk model *Time-Series* yaitu model deret waktu, kita harus berhati-hati, karena kegagalan model dapat menyebabkan identifikasi yang salah pada titik balik dalam deret tersebut. Prinsip objektivitas dan tahan uji mengarah pada beberapa prinsip metodologis untuk metode berbasis model. Khususnya, pembangunan model harus disertai dengan uji spesifikasi yang luas, untuk memastikan bahwa model tersebut kokoh. Penggunaan model harus dijelaskan secara eksplisit dan harus didokumentasikan dan dibuat transparan bagi pengguna.

Relevansi

Meskipun sumber *big data* mencakup seluruh populasi, informasi statistik yang dapat diekstraksi mungkin terbatas. Sebagai contoh, dengan sensor lalu lintas yang dilakukan oleh pihak kepolisian untuk menindak pelanggar (penjelasan kasus sebelumnya), kita dapat mengamati bahwa sebuah mobil sedang lewat, tetapi kita tidak tahu siapa yang ada di dalam mobil, siapa pemilik mobil tersebut, dan mengapa dia atau dia mengemudi di lokasi tersebut. Jadi, data ini dapat digunakan untuk statistik intensitas lalu lintas, tetapi relevansinya dan kohesinya dengan statistik lain tetap terbatas. (*Braaksma & Zeelenberg, 2020*)

Model dari Big Data dalam Modernisasi Official Statistic

Salah satu cara untuk mengimplementasikan *big data* dalam statistik resmi adalah dengan memandang agregat big data sebagai statistik di dalamnya. NSO dapat menerima *big data* apa adanya yaitu indikator dari *big data* yang tidak sempurna, namun sangat tepat waktu dan perkembangannya sangat pesat dalam masyarakat. NSO melaksanakan proses produksi statistik mereka dengan memanfaatkan berbagai metode yang beragam yaitu dari ranah statistik/matematika ataupun ranah ilmu komputer. Metode-metode tersebut telah berkembang di NSO dengan mengandalkan fitur-fitur khusus dari data yang dimaksudkan untuk mereka tangani. Oleh karena itu, metode-metode NSO saat

ini membawa jejak yang dapat dengan mudah dilacak kembali ke sifat data yang dikumpulkan dan dikonsumsi dalam konteks Statistik Resmi (selanjutnya akan disebut sebagai data tradisional).

Secara khusus, kualitas data tradisional dijamin oleh seperangkat metadata yang terdefinisi dengan baik. Dimensi waktu dari data tradisional sepenuhnya berada di bawah kendali masing-masing NSO, karena direncanakan dan dikelola sejak tahap desain awal survei. Sebagai produsen data tradisional, NSO tidak memiliki masalah aksesibilitas, dan bahkan saat menggunakan data dari arsip administratif, mereka dapat mengandalkan protokol pertukaran yang sudah mapan. Oleh karena itu, meskipun *Big Data* muncul sebagai peluang yang baik, kita harus mengakui bahwa metode NSO tradisional sepertinya belum siap untuk menangani kekhususan jenis data tersebut. Oleh karena itu, kemungkinan bagi NSO untuk menangani *Big Data* melalui dua cara: mengadopsi teknologi *Big Data* dan memikirkan kembali metode untuk memungkinkan analisis statistik yang kuat pada *Big Data*.

Upaya untuk memperkaya survei tradisional dengan informasi tambahan dari sumber *Big Data* akan memerlukan identifikasi komponen kecil dari sumber tersebut yang sebenarnya terhubung dengan unit dunia nyata yang diambil dalam survei. Ini akan memungkinkan untuk mentransfer informasi baru dari sumber *Big Data* ke unit sampel yang tepat. Dengan kata lain, untuk memanfaatkan peluang *Big Data* tanpa membahayakan kualitas inferensi, NSO harus dapat menghubungkan data survei dengan *Big Data*. (**Statistics Division Economic and Social Commission for Asia and the Pacific (ESCAP)**)

Isu Big Data

Walaupun big data berpotensi untuk menjadi salah satu sumber dan modernisasi dari official statistik, namun NSO perlu mencari solusi untuk isu-isu berikut

Output Statistik

Sumber *Big Data* dapat digunakan untuk menggantikan atau melengkapi sumber lain dalam produksi statistik yang merupakan bagian dari program statistik yang sudah ada. Beberapa contoh yang disebutkan sebelumnya memiliki potensi seperti penggunaan data deteksi *loop* lalu lintas untuk statistik lalu lintas, atau penggunaan pesan media sosial untuk menyusun indeks kepercayaan konsumen. Aplikasi-aplikasi media sosial tersebut dapat mengurangi beban administratif pada individu dan bisnis, atau mengurangi biaya lainnya. Jika hambatan-hambatan yang mungkin terkait dengan akses data, metodologi, privasi, dll., dapat diatasi, maka kasus bisnisnya menjadi jelas.

Sebenarnya, NSO harus menemukan optimum baru dan lebih kaya untuk output-nya, mengingat ketersediaan yang meningkat dari sumber data. Meskipun definisi *Big Data* merujuk pada sumber data, penting untuk berorientasi pada output dalam keputusan kebijakan. Penting juga untuk menyadari kebutuhan informasi dari pengguna statistik yang sebenarnya dan calon pengguna. Ada risiko bahwa hanya kebutuhan pelanggan saat ini yang dikenal oleh NSO. Terutama jika penempatan NSO menjadi masalah, kebutuhan eksternal harus diketahui.

Metodologi Statistik

Sebagian besar statistik bertujuan memberikan informasi tentang populasi individu atau bisnis atau set relevan lainnya, seperti barang impor atau terjual. Namun, populasi yang dicakup oleh *Big Data* mungkin tidak jelas. Tidak mengetahui komposisi populasi yang disertakan *dalam Big Data* menimbulkan pertanyaan tentang teori sampling apa yang harus diterapkan atau jika teori sampling tidak dapat diterapkan bagaimana selanjutnya? Pertimbangan kualitas umumnya penting saat memutuskan metode yang digunakan. Tetapi definisi kualitas memiliki masalahnya sendiri, terutama jika pendekatan pemodelan menjadi lebih dominan. Saat memutuskan masalah metodologis, aspek kualitas ini mungkin perlu perhatian khusus, terutama karena tidak berada dalam posisi untuk memberikan informasi tentang kualitas dan hubungan data dengan informasi lain yang tersedia secara publik.

Proses Dalam Statistik

Ketika menggunakan *Big Data*, desain proses statistik memerlukan perhatian khusus. Mungkin sulit untuk menerima dan memproses dataset dengan volume yang sangat besar, terutama jika "V" kedua dari *Big Data*, yaitu "kecepatan", berlaku. Penggunaan *Big Data* memerlukan penyiapan proses produksi yang lebih fleksibel, dengan waktu penyelesaian yang singkat. Bukan hanya pengumpulan data yang harus fleksibel, tetapi juga pemrosesan data lebih lanjut dalam rantai produksi. Secara umum, NSO yang mulai menggunakan *Big Data* mungkin harus beradaptasi atau bahkan mempertimbangkan ulang arsitektur perusahaannya.

Pertimbangan biaya tentu penting ketika merancang proses statistik. Dalam beberapa kasus, *Big Data* harus dibayar, meskipun NSO tidak terbiasa melakukan ini. Tetapi menggunakan *Big Data* juga dapat menghemat biaya, misalnya jika survei yang membutuhkan banyak tenaga kerja dapat digantikan oleh pengumpulan data yang sepenuhnya otomatis. Manfaat bagi penyedia data dalam bentuk pengurangan beban tanggapan dan bagi pengguna data dalam bentuk layanan statistik yang lebih baik bahkan mungkin menjadi pertimbangan yang lebih penting.

Privasi dan Keamanan

Ketika melihat privasi dan keamanan, ada beberapa masalah, nyata atau yang dirasakan, yang dapat menghambat penggunaan *Big Data*. Kepemilikan data dan hak cipta dapat menjadi masalah dan tujuan di mana data didaftarkan. Bahkan jika data dapat diakses secara publik, misalnya di situs web atau sebagai pesan media sosial yang tidak memiliki pembatasan akses, pertanyaan tentang kepemilikan dan tujuan publikasi dapat muncul.

Terdapat langkah-langkah yang dapat diambil oleh NSO untuk mengatasi setidaknya beberapa hambatan ini. Dalam beberapa kasus, penggunaan persetujuan informasi mungkin menjadi solusi. Jika NSO dapat menawarkan pengurangan beban tanggapan, hal ini dapat sangat membantu, juga dalam mendapatkan dukungan dari masyarakat umum. Pada tingkat hukum dan teknis saat ini, secara umum disarankan untuk tidak menyimpan data dan proses yang bersifat sensitif dan kritis di cloud. Pengguna layanan awan tetap bertanggung jawab atas keamanan dan privasi data, dan kepercayaan publik bergantung pada hal ini. Disarankan juga untuk menggunakan enkripsi data jika memungkinkan.

Isu lain yang akan Muncul di Masa Depan

Bagi NSO yang ingin menjadikan *Big Data* sebagai bagian serius dari bisnis statistik mereka, tata kelola dapat menjadi masalah. Karena aspek-aspek strategis yang penting dari *Big Data*, subjek ini seharusnya mendapatkan perhatian di tingkat manajemen tertinggi NSO. Menetapkan prioritas, menciptakan kondisi yang menguntungkan untuk menggunakan **Big Data**, dan mengambil keputusan anggaran terkait akan menjadi tugas untuk tingkat strategis, begitu juga dengan membuat pilihan kebijakan.

Masalah yang belum dibahas sejauh ini adalah modal manusia. Untuk bekerja dengan *Big Data*, diperlukan keterampilan teknis khusus, seperti keterampilan komputasi tingkat lanjut, pemahaman yang baik tentang matematika dan statistika, keterampilan pemodelan, dan keterampilan rekayasa data. Bekerja dengan *Big Data* membutuhkan pola pikir terbuka dan kemampuan untuk tidak melihat semua masalah secara *a priori* dalam istilah teori sampling. Cara untuk menangani masalah budaya ini adalah dengan membuat satu atau lebih pusat ilmuwan data yang bekerja dengan *Big Data*, dan biarkan pusat-pusat ini tumbuh, yang akan menjadi proses alami jika mereka berhasil. (Struijs & Daas, 2013)

SIMPULAN

Jelas bahwa *Big Data* akan memiliki dampak besar pada komunitas statistik dan statistik resmi. Khususnya dampak dari hal ini akan semakin jelas, tetapi beberapa fitur sudah terlihat atau dapat diprediksi. Jawaban atas pertanyaan-pertanyaan sebelumnya mengenai modernisasi, metodologi serta lainnya akan menjadi krusial terhadap dampak *Big Data* terhadap statistik resmi.

Secara umum, lembaga statistik nasional (NSO) seharusnya tidak takut untuk menggunakan model dalam menghasilkan statistik resmi, seperti yang tampaknya sudah mereka lakukan sebelumnya, asalkan model dan metode tersebut didokumentasikan dengan baik. Oleh karena itu, kita seharusnya lebih memperhatikan bagaimana model dapat digunakan untuk menghasilkan statistik resmi dari *big data*. Penting bagi NSO untuk terus menjelajahi peluang big data secara aktif. Banyak sumber lain akan muncul dan mungkin menjadi tersedia untuk produksi statistik. Di masa depan yang dekat, data biologis, misalnya, tentang genom, dan data medis, tentang kesehatan dan perawatan individu, akan menjadi tersedia untuk penelitian ilmiah dan untuk dikaitkan dengan data sosial tentang pendapatan, kejahatan, pekerjaan, dll. Eksplorasi peluang ini akan diiringi oleh tantangan yang tidak sederhana, seperti yang dijelaskan dalam bab ini, tetapi NSO berada dalam posisi yang sangat baik untuk menggunakan pengalaman tradisional dan standar kualitas tinggi mereka dengan cara yang inovatif.

REFERENSI

Braaksma, B., & Zeelenberg, K. (2020). *Big data in official statistics*.

- Letouzé, E., & Jütting, J. (2015). *OFFICIAL STATISTICS, BIG DATA AND HUMAN DEVELOPMENT ABOUT DATA-POP ALLIANCE*.
- Scannapieco, M., Virgillito, A., & Zardetto, D. (n.d.). *Placing Big Data in Official Statistics: A Big Challenge?*
- Statistics Division, E. (n.d.). *Using Big Data for Official Statistics: Key considerations when using mobile phone data*. <https://www.unescap.org/our-work/statistics>
- Struijs, P., Braaksma, B., & Daas, P. J. H. (2014). Official statistics and Big Data. In *Big Data and Society* (Vol. 1, Issue 1). SAGE Publications Ltd. <https://doi.org/10.1177/2053951714538417>
- Struijs, P., & Daas, P. (2013). *Topic (v): Integration and management of new data sources BIG DATA, BIG IMPACT?*
- West, B. T., Kreuter, F., & Jaenichen, U. (2013). “Interviewer” effects in face-to-face surveys: A function of sampling, measurement error, or nonresponse? *Journal of Official Statistics*, 29(2), 277–297. <https://doi.org/10.2478/jos-2013-0023>
- Zed, Mestika. 2008. *Metode Penelitian Kepustakaan*. Jakarta : Yayasan Obor Indonesia