

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/337911427>

Sains Data

Technical Report · November 2019

DOI: 10.13140/RG.2.2.21816.49924/2

CITATIONS

4

READS

12,544

1 author:



[Mahyuddin K. M. Nasution](#)

University of Sumatera Utara

401 PUBLICATIONS 3,607 CITATIONS

SEE PROFILE

Sains Data

Mahyuddin K M Nasution

Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Sumatera Utara, Padang Bulan
20155 USU, Medan, Indonesia¹

E-mail: mahyuddin@usu.ac.id

Abstract. Kehadiran sains baru tidak serta merta terjadi. Setiap sains dimulai dari kepentingan, pembicaraan, dan mencari landasan dasar, namun pada umumnya fondasi utama sains adalah matematika. Sains data meliputi pengetahuan terstruktur dan sistematis tentang data. Akan tetapi, banyak sains lain yang memiliki keterkaitan dengan data dimaksud, mulai dari statistik sampai ilmu komputer. Tulisan ini bertujuan untuk mengungkapkan hambatan dan keterbatasan sains itu menjadi sains data seutuhnya, atas dasar itu definisi sains data perlu diuraikan, kemudian meneguhkan sains data sebagai sains baru dan tidak tergantung langsung terhadap beberapa sains lain.

1. Pengantar

Sains data² merupakan gabungan dua kata yang membentuk istilah untuk merujuk kegiatan ilmiah disekitar atau berkaitan dengan apa yang dikenali dengan data [1], yaitu mulai dari pengumpulan dan pengolahannya, akhirnya menyajikannya sebagai informasi yang berguna bagi pengambilan keputusan atau bermanfaat kepada pihak-pihak yang berkepentingan dengan data [2, 3]. Sebagai sains, pembatasan tentangnya perlu diungkapkan, namun istilah rujukan itu tidak cukup menyatakan maksud dan tujuan keberadaannya sebagai sains, yang menyebabkan berbagai definisi tentangnya bermunculan [4, 5]. Suatu sains, memiliki dasar secara ontologi, dan secara taksonomi menyebar ke berbagai arah perkembangan, namun masih berada dalam ruang lingkup yang saling terkait [6, 7].

Sains data melibatkan metode dalam semua kegiatan atau keilmuannya [8], namun terpetakan secara logis dalam satu integrasi keilmiah [9, 10]. Pada sisi lain, data secara khusus merupakan objek kajian yang telah lama menjadi bagian dari keilmuan-keilmuan lain secara berbeda [11], sedangkan sains lahir secara teoritis untuk mengantarkan teknologi dan penerapan lainnya yang merintis peningkatan mutu hidup manusia [12, 13]. Oleh karena itu, sains data menjadi sistem paradigma yang melibatkan empiris, teoritis, komputasi, dan data raksasa.

2. Ulasan berdasarkan jejak

Sains data sebagai sistem keilmuan merupakan sistem terbuka [14], yang terdiri dari satuan-satuan yang saling berinteraksi. Dengan demikian, satuan-satuan dalaman berinteraksi dengan satuan-satuan eksternal [15]. Satuan dalaman fokus kepada data, sedangkan satuan eksternal

¹ Sains Data 1(1) 2019, Program Studi "Data Science & Artificial Intelligence", Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Sumatera Utara, Sumatera Utara, Indonesia

² Tulisan ini merupakan oleh-oleh perjalanan tugas kunjungan kajian tentang *Data Science & Artificial Intelligence* (DS&DI) bersama tim Universitas Sumatera Utara, Athens - Leiden, November 2019

melengkapi dari luar. Apabila sains data adalah sebagai sains tentang data (atau keilmuan yang terkait dengan data secara keseluruhan), maka sains data sebagai sistem memerlukan keseluruhan satuan itu terorganisir didalam suatu struktur dan sistematika keilmuan [16].

Data adalah sesuatu yang fenomenal saat ini, yang memerlukan wadah kajian secara khusus. Walaupun data telah lama menjadi bagian utama dari statistika. Berdasarkan matematika, statistika tidak memiliki kemampuan lebih tanpa melibatkan konsep intinya, yaitu probabilitas [17, 18], walaupun statistika sudah berkududukan mapan dengan analisis data [19, 20]. Statistik baik secara teori maupun penerapan terdiri dari kombinasi keterlibatan jumlah Σ sampai rata-rata μ , atau sentralitas data dan beberapa perluasan mereka [21]. Statistika belum tidak mampu memprediksi, menguji dan menilai sesuatu, sebelum membuktikan rumusan itu dalam kewajaran menurut kontur probabilitas [22, 23]. Banyak ilmuwan menyatakan bahwa ilmu data menurut istilah pertama tidak lain adalah statistika itu sendiri [24]. Namun, data sebagai fenomena tidak dapat diungkapkan oleh statistika secara tuntas. Statistika secara umum menyasar kuantitas atau kualitas dengan wadah parameterik dan non-perametrik. Akan tetapi, data memiliki sistematika struktur tersendiri di samping konsep taburan [25, 26].

Turunan dari data, atau kadangkala sebagai saudara kembar data, atau juga pada saat tertentu sebagai sumber data, yaitu informasi, menjadi bagian terpenting dari semua aktivitas manusia saat ini [27, 28]. Dalam hal ini, metode memiliki peranan, untuk menyelesaikan persoalan-persoalan yang terpaut dengan atau dari/ ke data itu. Secara komputasi, statistika tumbuh dalam teori untuk menghadirkan metode. Namun, implikasi metode itu hanya implementasi rumus-rumus statistika didalam alat hitung, seperti komputer dan kelengkapan lainnya [29, 30]. Statistika dalam teorinya tidak memiliki cara untuk menggunakan kemudahan secara tepat guna dan berdaya guna [31]. Padahal, kemudahan komputasi seperti memori dan kecepatan prosesor memerlukan keseimbangan kinerja di antara mereka.

Syarat utama dari kontur statistik, yaitu probabilitas, adalah keacakan [32]. Untuk mengabaikan pengujian, banyak data minimal sebagai sampel oleh alasan statistik dipasang sebagai asumsi awal pengolahan data dan memasang pembanding rujukan [33], sebagaimana dataset [34]. Alih-alih mencari cara menguji keabsahan sampel dari populasi yang berlaku umum, memandang pula sains data = statistika [35]. Ini berarti menganggap data = komputasi = rumus [36, 37], yang secara matematika tidak memiliki kesimpulan abstraksi yang belum sah terbukti. Ini mengakibatkan penambangan data, sebagai pergeseran pemahaman terhadap data, tanpa sadar telah mengukudeta tata aturan pengolahan data, sehingga tidak jarang untuk data yang sama terdapat kesimpulan yang berbeda [38].

Apakah itu sekedar rekognisi atau sebagai proses sebelum data [39]. Apakah itu presentasi atau sekedar sajian luaran. Jelasnya, terdapat rangkaian kegiatan ilmiah sebelum dan sesudah data wujud. Latar belakang data, siapa pengumpul atau asal usulnya, akan menyebabkan informasi sebagai hasil olahan data memungkinkannya tidak sah sebagai sumber pengetahuan [40, 41]. Perlukah melakukan forensik? [42, 43] Jawabannya terletak pada menempatkan posisi statistika dalam rangkaian keilmuan itu [44]. Tidak sedikit bidang keilmuan yang bergumul dengan data, statistika sebagian darinya. Bagaimanapun, dasar semua itu adalah matematika, yang pada awalnya terbagi ke dalam empat bidang utama: Aritmatika, Aljabar, Trigonometri, dan Geometri. Dengan dasar itu, bagaimanapun statistika berusaha melepaskan diri dari jeratan aritmatika, perangkat komputasi menempatkan statistika tumbuh sekitar ilmu hitung yang melibatkan angka dan operator [45]. Seiring dengan itu, tuntutan terhadap pemaknaan data meminta kehadiran bidang lain seperti optimasi, matrik, jarak dan similaritas, penelitian operasi, dan lainnya [46]. Bukankah, teori kabur dan himpunan kasar juga sebagai akibat tuntutan kepentingan terhadap data [47].

3. Suatu pendekatan

Kelahiran suatu sains dimulai dengan munculnya istilah itu dalam kalangan akademisi tertentu [48]. Lalu eksistensi sains itu terbangun ketika menguraikannya melahirkan banyak dokumen yang terbit.

Istilah-istilah terkait akan hadir mengimbangi istilah yang mungkin menjadi nama dari sains itu. Sains data adalah suatu istilah, dan definisi-definisi tentangnya hadir bersama dokumen yang menguraikannya. Saat ini, uraian itu hadir bersama ruang informasi sebagai hasil diskusi dan tukar pendapat. Terdapat dua ruang berbeda yang menjadi tumpuan perhatian, pertama ruang informasi dari mesin cari di mana terdapat lalu lalang informasi yang terekam sebagai unjuk kepentingan dari beragam pihak yang berkepentingan tentangnya [49]. Istilah sains data direkam dan diungkap berdasarkan tahun diperbincangkan, untuk melihat naik turunnya aktivitas kegiatan tentang sains data. Tentunya secara semantik, informasi ini melibatkan seluruh pengguna internet yang terkoneksi dan merekamkan kepentingannya dengan tanggapannya [50].

Selain itu, sebagai penyeimbang terhadap kepentingan itu, dihadirkan informasi tentang dokumen-dokumen terkait dengan istilah sains data yang terindeks oleh mesin cari dalam bentuk jumlah dan tahun [51]. Dengan alasan bahwa dokumen-dokumen itu menjadi informasi yang dapat dipercaya karena berasal dari para ilmuwan yang terkait dengan beragam bidang ilmu sebagai pemangku kepentingan terhadap wujudnya ilmu baru. Gerak grafik akan menggambarkan lintasan perjalanan suatu keilmuan.

4. Pendirian dalam diskusi

Untuk mendirikan suatu ilmu baru, istilah terus mengalir dalam setiap kegiatan ilmiah terkait. Definisi-definisi mengisi ruang diskusi, dan kemudian menetapkan, walaupun suatu definisi hanya berlaku sepanjang tidak ditemukan keberatan.

4.1. Istilah

Istilah sains menyiratkan pengetahuan yang tersusun secara sistematis dan terstruktur [52]. Sistematis merujuk kepada suatu usaha atau kajian untuk membangun dan mengatur pengetahuan dalam bentuk penjelasan sehingga menghasilkan suatu batasan sebagaimana hadirnya definisi di dalam sains itu, kemudian disusun dengan teori berupa dalil dan bukti [16]. Semua itu tersusun secara nalar dan logis, dan bagian-bagian itu menjadi suatu struktur dari sains tersebut. Berdasarkan itu, semula sains data mengalami ilmu komputer untuk mendasari keilmuannya [54], namun jebakan kompleksitas yang tidak terselesaikan pada algoritma [55, 56], dengan mana itu sebagai fokus ilmu komputer, telah menyebabkan ilmu komputer mengalami defraksi keilmuan, dan tidak dapat sebagai landasan kuat suatu keilmuan baru.

Sebaliknya, sains komputer tidaklah untuk menggantikan sains data. Kegagalan penegasan sains data itu menghadirkan ide untuk memunculkan istilah datalogi [57, 58]. Berdasarkan sifat dan karakteristik data dalam dimensinya, data menjadi bagian dari keseluruhan keilmuan yang ada [59, 60]. Sebagaimana kata logi yang mencoba menegaskan kata metode dalam metodologi [61], istilah datalogi dengan demikian cenderung menjadikannya sebagai bagian dari setiap sains yang ada dengan mana data menjadi pendukung sebarang kajian di dalamnya [62]. Oleh karena itu, datalogi semakin tidak mempertegas keberadaan ilmu yang dimaksudkan sebagai ilmu data.

Ilmu data sebagai istilah adalah untuk menyatakan data dan apa yang terdapat di sekitarnya, mulai dari keberadaannya dan maknanya. Namun, studi sistematis tentang data - organisasi, properti, dan analisis, atau perannya dalam inferensi - di dalam statistika memberi kekangan jika istilah sains data menggantikan statistika. Fakta menyatakan bahwa sekarang ini, berbicara tentang data itu berarti bahwa berhadapan dengan sejumlah besar data, atau dikenali dengan data raksasa [63], yang menyebabkan beberapa konsep statistik perlu mengalami perubahan [64]. Pengorganisasian data tidak terbatas pada sekedar angka-angka. Karakteristik data

berlimpah ruah sepanjang terkait dengan pemaknaan kehidupan, atribut data tidak seperti properti yang tersedia dalam statistika [65]. Analisis data oleh statistika terkekang oleh sampel, dan mengalami ganjalan bila berhadapan dengan ketidakjelasan yang dipamerkan data raksasa (antara sebagai sampel atau populasi) [66, 67]. Dengan kata lain, statistika berurusan dengan konvergensi dalam komputasi [68], dan dengan itu istilah sains data adalah jalan damai yang berkaitan dengan penguraian suatu disiplin yang secara khusus melibatkan campuran statistika dan penghitungan berskala besar [69]. Oleh karena itu, istilah sains data sebagai frase dalam hal ini menjadi penegasan terhadap tugas baru berkaitan dengan data [70].

Selain istilah sains komputer dan statistika, tidak sedikit hadir istilah lain sebagai nama ilmu seperti penambangan data untuk mengaburkan pentingnya sains data [38]. Dalam pandangan penambangan data, anjungan-anjungan pengeboran akan menambang data raksasa seperti tambang minyak dalam kolam yang sama. Walaupun, data raksasa terkotak-kotak dalam sistem-sistem yang telah ada, tetapi sebenarnya data raksasa berada dalam ruang informasi yang tidak memiliki struktur apapun. Dengan demikian, penambangan data sesuai dengan metodenya hanya mampu mengungkap sebagian demi sebagian dari keseluruhan data raksasa [71, 72, 73]. Istilah sains data tidaklah penambangan data dalam arti keseluruhannya, atau sebaliknya.

4.2. Pendefinision

Sebagai sains baru, sains data mendapatkan penguatan dengan adanya diseminasi yang telah dilakukan oleh ilmuwan atau organisasi baik melalui kuliah atau pertemuan ilmiah [74]. Perbincangan tentang sains data menjadi isu utama dalam dunia ilmiah dengan hadirnya jurnal-jurnal yang bertugas sebagai sarana terbit artikel-artikel terkait dengan riset atau ulasan tentang keilmuan ini, meliputi ruang lingkup kajian yang mungkin hadir [75]. Seiring dengan itu, definisi baru tentang sains data dan tambahan ruang lingkup kajian disajikan untuk menjelaskannya.

Sekali lagi, konsep dasar statistika yang tidak pernah keluar dari jebakan aritmatika dalam selang $[0, 1]$, dari teori sampai komputasinya [76], diikuti oleh ilmu komputer yang harus puas berada dalam kajian dan penerapan "program = struktur data + algoritma" dengan jebakan kompleksitas [77]. Statistika tidaklah perlu berubah menjadi sains data, sedangkan ilmu komputer tetap harus mampu menegaskan diri dalam keilmuannya. Para statistikawan hendaklah mengurai hambatan konvergensi agar mampu menangani transformasi data, dari klasik menjadi kabur atau kasar (himpunan kasar). Lahirnya bidang lain yang mengkaji dari sudut lain tentang data dan komputer telah memilah-milah beberapa turunan keilmuan atau bidang. Fokus kajian dari ilmu komputer jadi berbeda dengan sistem/sains informasi, teknologi informasi maupun sistem komputer, misalnya. Walaupun, terdapat ilmuwan yang menyatakan bahwa "Data science is the child of statistics and computer science. [78].", secara teori sains data adalah sains tentang data. Perbedaan terjadi hanya sebagai akibat dari pengorganisasian satuan-satuan keilmuan yang diperlukan ke dalam satu sistem keilmuan untuk menghadapi dimensi data. Sistem ini didasarkan atas interaksi dan kesinambungan satuan-satuan keilmuan itu. Walaupun, semua keilmuan itu mendasarkan diri kepada matematika diskrit sebagai penggerak energi keilmuan, tetapi akan memberikan implikasi berbeda ketika penafsiran dilakukan berdasarkan misi dan visi keilmuan serta target luaran. Dengan demikian, sains apapun yang akan lahir, akan menyandarkan diri kepada matematika sebagai landasan keilmuan.

Sains data terdiri dari satuan-satuan keilmuan yang terorganisir secara terbuka, tetapi memiliki batasan tersendiri. Batasan dimaksudkan untuk membatasi kajian sesuai dengan target luaran dan sasaran capaian, tetapi juga tetap terbuka untuk mengenali perubahan-perubahan yang diperlukan. Ada tujuan yang tidak tegas, bahwa para saintis ingin menghasilkan sesuatu agar orang bisa menilai [79]. Substansi dari sains data berasal dari beragam keilmuan atau melibatkan investigasi multidisiplin, dan didukung penerapan teknologi. Bagaimanapun model data menjadi tumpuan dari investigasi, tetapi untuk membangun model itu memerlukan

rekognisi data secara keseluruhan. Model data mengusulkan ruang gerak pilihan metode untuk mengakses data agar analisis data memiliki kemampuan untuk bertumpu dalam menghasilkan informasi. Kolaborasi statistika, optimasi, dan metode penambangan, termasuk inferensi probabilistik adalah suatu pilihan yang menarik agar pengetahuan hadir [44]. Dengan berbagai kendala yang dimiliki metode itu, kecerdasan buatan hadir secara terintegrasi. Dengan demikian, sains data tidak saja berkaitan dengan satuan-satuan dari rekognisi data atau model data, statistika, optimasi, penambangan data, kecerdasan buatan, tetapi melibatkan dukungan dari teknologi yang tersedia dalam bentuk komputasi dengan segala perangkatnya (sistem perangkat keras, sistem perangkat lunak, dan algoritma) [80, 81], tetapi tidak menjadikan mereka sebagai substansi kajian [82].

Sebagaimana data dengan segala karakteristiknya, sains data sangat terkait erat dengan semua konsep penting lainnya tentang data itu sendiri seperti data raksasa dan pengambilan keputusan. Penegasan ini juga memberikan sesuatu komitmen terhadap data. Perekaman data hendaklah melibatkan validasi yang baik, forensik tentang asal usul data menjadi bagian pengkoleksian data secara cerdas [83], karena bagaimanapun prinsip penggunaan teknologi masih berlaku, yaitu *"garbage in garbage out"* (GIGO), dengan mana data perilaku dan ekonomi memiliki sifat berbeda, dengan mana terdapat subjektif mereka terhadap sistem [84].

4.3. Jejak rekam

Jejak rekam tentang perkembangan sains data sebagai sains baru terlihat dari pertumbuhan informasi dalam ruang informasi. Informasi terkait dengan sains data diungkap dari mesin cari Google mulai dari tahun pertama istilah ini hadir dalam literatur, dan ditegaskan melalui kajian-kajian dengan bukti dokumen terindeks pada Google Scholar. Penggunaan istilah sains data sebagai nama terhadap sains baru ini telah mencapai titik kulminasi, dan para saintis mengkaji kelengkapannya dalam tajuk-tajuk berbeda. Ini ditunjukkan oleh menurunnya persentase jumlah dokumen terkait dengan sains data dibandingkan dengan informasi tentang situs-situs kelompok riset atau kelengkapan lainnya yang terkait dengan sains data [85].

Perdebatan tentang apa itu sains data telah berakhir, dan sains data diterima sebagai sains baru yang terkait secara keseluruhan dengan data, berbeda dengan statistika, ilmu komputer, penambangan data, dan lainnya.

4.4. Definisi

Berhadapan dengan data, persoalan adalah secara utuh untuk menyatakan sains data, perlu mempertimbangkan rangkaian keterkaitan: data (δ), informasi (ι), dan pengetahuan (κ), dalam suatu hubungan, seperti dinyatakan sebagai berikut [86]: *"Data science is the extraction of knowledge from high-volume data, using skills in computing science, statistics and the specialist domain knowledge of experts."* Akan tetapi, ketika ekstraksi sesuatu dilakukan dari sumbernya, misalkan Ω menyatakan sumber, γ menyatakan fungsi ekstraksi yang melibatkan kecerdasan buatan [87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97], sains data adalah

$$DS(\delta, \iota, \kappa) = \gamma(\Omega) + \mu(\Sigma)$$

dengan mana μ adalah fungsi melibatkan perangkat yang tersedia melalui pengetahuan lain Σ .

5. Kesimpulan

Secara khusus, sains data telah distimulus oleh berbagai pakar terkait. Berbagai istilah dimunculkan untuk memberikan sumbang-saran dan peransang, berbagai tanggapan dari tergambar dari masyarakat maupun ilmuwan lain dengan hadirnya dokumen kajian yang tersaji dalam berbagai kegiatan ilmiah. Sebagai penelusuran, sains data adalah sains baru walaupun data telah lama dikenali dalam semua aktivitas ilmiah. Selanjutnya, berdasarkan kepentingan

penetapan sains data itu perlu pengkajian tentang status terkini dari sains data secara kronologi. Furthermore, based on the importance of data science determination, it is necessary to study the current status of chronological data science.

References

- [1] L Manovich 2015 Data science and digital art history *International Journal for Digital Art History* **1**.
- [2] E K Nwabueze, P Ranch 2005 Methods for dynamically accessing, processing, and presenting data acquired from disparate data sources *Unites States Patent* No:USOO6959306B2.
- [3] P Obrador 2006 Presenting a collection of media objects *Unites States Patent* No:US007149755B2.
- [4] M K M Nasution 2007 SumutSiana *Renungan*, IPR: EC00201944654. DOI:10.13140/RG.2.2.10127.59047.
- [5] M K M Nasution 2018 SumutSiana *IOP Conference Series: Materials Science and Engineering* **309(1)**. DOI:10.1088/1757-899X/309/1/012131.
- [6] M K M Nasution 2017 Ontologi *Ontologi dan Taksonomi Informasi* **1**, IPR:EC00201945521. DOI:10.13140/RG.2.2.22463.92323.
- [7] M K M Nasution 2018 Ontology *Journal of Physics: Conference Series* **1116(2)**. DOI:10.1088/1742-6596/1116/2/022030.
- [8] D Donoho 2017 50 Years of data science *Journal of Computational and Graphical Statistics* **26(4)**.
- [9] S Iwata 2008 Editor's Note: Scientific "Agenda" of data science *Data Science Journal* **7**.
- [10] C A Mattmann 2013 A vision for data science *Nature* **493**.
- [11] F Xia, W Wang, T M Bekele, H Liu 2017 Big scholarly data: A survey *IEEE Transactions on Big Data* **3(1)**.
- [12] M Bunge 1975 What is a quality of life indicator? *Social Indicator Research* **2**.
- [13] M K M Nasution 2019 Kelahiran sains *Translasi* **4**, IPR:EC00201982905. DOI:10.13140/RG.2.2.17622.80965/1.
- [14] K S Baker, G C Bowker 2007 Information ecology: open system environment for data, memories, and knowing *Journal of Intelligent Information System* **29**.
- [15] L Berchicci 2013 Towards an open R&D system: Internal R&D investment, external knowledge acquisition and innovative performance *Research Policy* **42(1)**. DOI:10.1016/j.respol.2012.04.017.
- [16] M K M Nasution, I Aulia, M Elveny 2019 Data *Journal of Physics: Conference Series* **1235(1)**. DOI:10.1088/1742-6596/1235/1/012110.
- [17] G Shafer 1990 The unity and diversity of probability *Statistical Science* **5(4)**.
- [18] M Borovcnik 2011 Strengthening the role of probability within statistics curricula *Teaching Statistics in School Mathematics-Challenges for Teaching and Teacher Education*, NISS **14**.
- [19] J O Ramsay, B W Silverman 2005 *Functional Data Analysis* Springer.
- [20] P Mihas 2019 Qualitative data analysis *Oxford Research Encyclopedias*. DOI:10.1093/acrefore/9780190264093.013.1195.
- [21] D G Altman, J M Bland 1995 Statistics notes: The normal distribution *BMJ* doi:10.1136/bmj.310.6975.298.
- [22] T Asparouhov, B Muthen 2006 Robust chi square difference testing with mean and variance adjusted test statistics *Mplus Web Notes* **10**.
- [23] K-H Yuan, P M Bentler 2011 Normal theory based test statistics in structural equation modelling *British Journal of Mathematical and Statistical Psychology* **51(2)**. DOI:10.1111/j.2044-8317.1998.tb00682.x.
- [24] W J Youden 1950 Index for rating diagnostic tests *Cancer* **3(1)**. DOI:10.1002/1097-0142(1950)3.
- [25] J Elith, M A Burgman, H M Regan 2002 Mapping epistemic uncertainties and vague concepts in predictions of species distribution *Ecological Modelling* **157(2-3)**.
- [26] W Chanhom, C Anutariya 2019 TOMS: A linked open data system for collaboration and distribution of cultural heritage artifact collections of National Museums in Thailand *New Generation Computing* **37(4)**. DOI:10.1007/s00354-019-00063-1.
- [27] S Rapps, E J Weyuker 1985 Selecting software test data using data flow information *IEEE Transactions on Software Engineering* **SE-11(4)**.
- [28] M Chen, D Ebert, H Hagen, R S Laramée, R van Liere, K-L Ma, W Ribarsky, G Scheuermann, D Silver 2009 Data, Information, and Knowledge in Visualization *IEEE Computer Graphics and Applications* **29(1)**. DOI:10.1109/MCG.2009.6.
- [29] K H Yuan 2005 Fit indices versus test statistics *Multivariate behavioral research* **40(1)**. DOI:10.1207/s15327906mbr4001_5.
- [30] N Siripon 2013 A novel design of distributed oscillator based on the balanced oscillator technique *2013 10th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology, ECTI-CON 2013*. DOI:10.1109/ECTICon.2013.6559633.

- [31] J L Bentley, M I Shamos 1978 A problem in multivariate statistics: algorithm, data structure and applications *CARNEGIE-MELLON UNIV PITTSBURGH PA DEPT OF COMPUTER SCIENCE*.
- [32] P M Bentler, K-H Yuan 1999 Structural equation modeling with small samples: Test statistics *Multivariate Behavioral Research* **34(2)**.
- [33] P J Fleming, J J Wallace 1986 How not to lie with statistics: The correct way to summarize benchmark results *Communication of the ACM* **29(3)**.
- [34] A Moore, M S Lee 1998 Cached sufficient statistics for efficient machine learning with large datasets *Journal of Artificial Intelligence Research* **8**.
- [35] N B-n Vilches, L Napalkova 2016 Application of data science techniques to the field of air traffic control *Universitat Autònoma de Barcelona*.
- [36] O Rosenbach 1953 A contribution to the computation of the "second derivative" from gravity data *Geophysics* **18(4)**. DOI:10.1190/1.1437943.
- [37] K A Berteussen, B Ursin 1983 Approximate computation of the acoustic impedance from seismic data *Geophysics* **48(10)**. DOI:10.1190/1.1441415.
- [38] W van der Aalst 2016 Data science in action *Process mining*.
- [39] I Fischler, P A Bloom 1979 Automatic and attentional processes in the effects of sentence contexts on word recognition *Journal of Verbal Learning and Verbal Behavior* **18(1)**.
- [40] M K M Nasution, M Hardi, R Sitepu 2016 Using social networks to assess forensic of negative issues *Proceedings of 2016 4th International Conference on Cyber and IT Service Management, CITSM 2016*. DOI:10.1109/CITSM.2016.7577513.
- [41] M K M Nasution, M Hardi, R Sitepu, E Sinulingga 2017 A Method to Extract the Forensic about Negative Issues from Web *IOP Conference Series: Materials Science and Engineering* **180(1)**. DOI:10.1088/1757-899X/180/1/012241.
- [42] M K M Nasution, D Sitompul, M Harahap 2018 Modeling reliability measurement of interface on information system: Towards the forensic of rules *IOP Conference Series: Materials Science and Engineering* **308(1)**. DOI:10.1088/1757-899X/308/1/012042.
- [43] M K M Nasution 2019 Forensic in information technology: A redefinition *Journal of Physics: Conference Series* **1235(1)**. DOI:10.1088/1742-6596/1235/1/012106.
- [44] W S Cleveland 2001 Data science: An action plan for expanding the technical areas of the field of statistics *International Statistical Review* **69**.
- [45] J B Kruskal 1977 Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics *Linear Algebra and its Applications* **18(2)**.
- [46] W Noor, M N Dailey, P Haddawy 2014 Learning predictive choice models for decision optimization *IEEE Transactions on Knowledge and Data Engineering* **26(8)**. DOI:10.1109/TKDE.2013.173.
- [47] M K M Nasution 2018 The uncertainty: A history in Mathematics *Journal of Physics: Conference Series* **1116(2)**. DOI:10.1088/1742-6596/1116/2/022031.
- [48] M K M Nasution 2020 The birth of a science *IOP Conference Series: Earth and Environmental Science*.
- [49] M K M Nasution 2018 Singleton: A role of the search engine to reveal the existence of something in information space *IOP Conference Series: Materials Science and Engineering* **420(1)**. DOI:10.1088/1757-899X/420/1/012137.
- [50] M K M Nasution 2018 Semantic interpretation of search engine resultant *IOP Conference Series: Materials Science and Engineering* **300(1)**. DOI:10.1088/1757-899X/300/1/012053.
- [51] M K M Nasution 2018 Doubleton: A role of the search engine to reveal the existence of relation in information space *IOP Conference Series: Materials Science and Engineering* **420(1)**. DOI:10.1088/1757-899X/420/1/012138.
- [52] J L Heilbron 2003 *The Oxford Companion to the History of Modern Science* Oxford University Press.
- [53] P M Dung, G Sartor 2011 The modular logic of private international law *Artificial Intelligence and Law* **19(2-3)**. DOI:10.1007/s10506-011-9112-5.
- [54] H A Gohel 2015 Data science - data, tools & technologies *CSI Communication* **8**.
- [55] V Cutello, G Nicosia, M Pavone 2004 Exploring the Capability of Immune Algorithms: A characterization of hypermutation operators *International Conference on Artificial Immune Systems LNCS* **3239**. DOI:10.1007/978-3-540-30220-9_22.
- [56] D I De Silva, N Kodagoda, S R Kodituwakku, A J Pinidiyaarachchi 2017 Analysis and enhancements of a cognitive based complexity measure *IEEE International Symposium on Information Theory - Proceedings*. DOI:10.1109/ISIT.2017.8006526.
- [57] P Naur 1966 The science of datalogy *Communication of the ACM*.
- [58] E Sveinsdottir, E Frokjaer 1988 Datalogy - The copenhagen tradition of computer science *BIT Numerical Mathematics* **28(3)**.
- [59] C A Chinn, W F Brewer 1993 The role of anomalous data in knowledge acquisition: A theoretical framework

and implication for science instruction *Center for the Study of Reading* Technical Report No. 583.

- [60] L M Ghiringhelli, J Vybiral, S V Levchenko, C Draxl, M Scheffer 2015 Big data of materials science: Critical role of the descriptor *Physical Review Letters* **114**.
- [61] M K M Nasution 2019 Metodologi *Translasi* **10**. DOI:10.13140/RG.2.2.17349.60646.
- [62] M K M Nasution 2019 Research methodology *IOP Conference Series: Materials Science and Engineering*
- [63] L Manovich 2015 Data science and digital art history *DAH-Journal* **1**.
- [64] F Provost, T Fawcett 2013 Data science and its relationship to big data and data-driven decision making *Big data* **1(1)**.
- [65] W R van Hage, M van Erp, V Malaisé 2012 Linked Open Piracy: A Story about e-Science, Linked Data, and Statistics *Journal on Data Semantics* **1(3)**.
- [66] A Gandomi, M Haidar 2015 Beyond the hype: Big data concepts, methods, and analytics *International Journal of Information Management* **32(2)**. DOI:10.1016/j.ijinfomgt.2014.10.007.
- [67] G Grolemond, H Wickham 2018 R for data science *Journal of Statistical Software* **77(1)**.
- [68] M K M Nasution 2017 Modelling and simulation of search engine *Journal of Physics: Conference Series* **801(1)**. DOI:10.1088/1742-6596/801/1/012078.
- [69] J B Greenhouse 2013 *Statistical Thinking: The bedrock of data science* The Huffington Post.
- [70] J Hardin, R Hoerl, N J Horton, D Nolan, B Baumer, O Hall-Holt, P Murrell, R Peng, P Roback, D T Lang, M D Ward 2015 Data science in statistics curricula: Preparing students to "think with data" *The American Statistician* **69(4)**.
- [71] M K M Nasution 2016 Social network mining (SNM): A definition of relation between the resources and SNA *International Journal on Advanced Science, Engineering and Information Technology* **6(6)**. DOI:10.18517/ijaseit.6.6.1390.
- [72] M K M Nasution, M Hardi, R Syah 2017 Mining of the social network extraction *Journal of Physics: Conference Series* **801(1)**. DOI:10.1088/1742-6596/801/1/012020.
- [73] M K M Nasution 2019 Social Network Mining: A discussion *Journal of Physics: Conference Series* **1235(1)**. DOI:10.1088/1742-6596/1235/1/012111.
- [74] M K M Nasution 2018 Indonesia knowledge dissemination: A snapshot *Journal of Physics: Conference Series* **978(1)**. DOI:10.1088/1742-6596/978/1/012012.
- [75] P F Uhler, P Schroder 2007 Open data for global science *Data Science Journal* **6**.
- [76] S R Kalidindi, M De Graef 2015 Materials data science: Current status and future outlook *Annu. Rev. Mater. Res.* **45**.
- [77] X Liu, D Li, S Wang, Z Tao 2007 Effective algorithm for detecting community structure in complex networks based on GA and clustering *International Conference on Computational Science, Computational Science ICCS 2007* DOI:10.1007/978-3-540-72586-2_95.
- [78] D M Blei, P Smyth 2017 Science and data science *PNAS* **114(33)**.
- [79] D J Patil 2011 Bulding data science teams *O'Reilly*.
- [80] T F Abidin, R Ferdhiana 2017 Algorithm for updating n-grams word dictionary for web classification *2016 International Conference on Informatics and Computing, ICIC 2016*. DOI:10.1109/IAC.2016.7905758.
- [81] S Xiang, H Zhu, X Wu, L Xiao, M Bonsangue, W Xie, L Zhang 2020 Modeling and verifying the topology discovery mechanism of OpenFlow controllers in software-defined networks using process algebra *Science of Computer Programming* **187**. DOI:10.1016/j.scico.2019.102343.
- [82] M Molina-Solana, M Ros, M D Ruiz, J Gomez-Romero, M J Martin-Bautista 2017 Data science for building energy management: A review *Renewable and Sustainable Energy Reviews* **70**.
- [83] M K M Nasution 2019 Konsep forensik data *Translasi* **1** IPR:EC00201981615. DOI:10.13140/RG.2.2.35668.96643.
- [84] S Chawla, J Hartline, D Nekipelov 2014 Mechanism design for data science arXiv:1404.5971v2 [cs.GT].
- [85] M K M Nasution, R Syah, M Elveny 2017 Studies on behaviour of information to extract the meaning behind the behaviour *Journal of Physics: Conference Series* **801(1)**. DOI:10.1088/1742-6596/801/1/012022.
- [86] V Dhar 2012 Data science and prediction *Commun ACM* **56**.
- [87] M K M Nasution, S A Noah 2010 Superficial method for extracting social network for academics using web snippets *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **6401 LNAI**. DOI:10.1007/978-3-642-16248-0_68.
- [88] M K M Nasution, S A Noah 2011 Extraction of academic social network from online database *2011 International Conference on Semantic Technology and Information Retrieval STAIR 2011*. DOI:10.1109/STAIR.2011.5995766.
- [89] M K M Nasution, S A Noah 2012 Information retrieval model: A social network extraction perspective *Proceedings - 2012 International Conference on Information Retrieval and Knowledge Management, CAMP'12*. DOI:10.1109/InfRKM.2012.6204999.
- [90] M K M Nasution 2014 New method for extracting keyword for the social actor *Lecture Notes in Computer*

Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)
8397 LNAI(PART 1). DOI:10.1007/978-3-319-05476-6_9.

- [91] M K M Nasution, O S Sitompul 2017 Enhancing extraction method for aggregating strength relation between social actors *Advances in Intelligent Systems and Computing* **573**. DOI:10.1007/978-3-319-57261-1_31.
- [92] M K M Nasution, S A Noah 2017 Social Network Extraction Based on Web. A Comparison of Superficial Methods *Procedia Computer Science* **124**. DOI:10.1016/j.procs.2017.12.133.
- [93] C Anutariya, R Dangol 2018 VizLOD: Schema Extraction and Visualization of Linked Open Data *Proceeding of 2018 15th International Joint Conference on Computer Science and Software Engineering, JCSSE 2018*. DOI:10.1109/JCSSE.2018.8457325.
- [94] M Elfida, M K M Nasution, O S Sitompul 2018 Enhancing to method for extracting Social network by the relation existence *IOP Conference Series: Materials Science and Engineering* **300(1)**. DOI:10.1088/1757-899X/300/1/012057.
- [95] M K M Nasution 2018 Social network extraction based on Web: 1. Related superficial methods *IOP Conference Series: Materials Science and Engineering* **300(1)**. DOI:10.1088/1757-899X/300/1/012056.
- [96] M K M Nasution 2018 Social network extraction based on Web: 2. Strategies in superficial methods *Journal of Physics: Conference Series*. DOI:10.1088/1742-6596/1116/2/022029.
- [97] M K M Nasution, O S Sitompul, S A Noah 2018 Social network extraction based on Web: 3. The integrated superficial method *Journal of Physics: Conference Series* **978(1)**. DOI:10.1088/1742-6596/978/1/012033.