PROTEKSI PRIVASI BIG DATA DALAM MEDIA SOSIAL

Winarsih dan Irwansyah
Universitas Indonesia
uiwinarsih@gmail.com

Abstract

Perkembangan media sosial di Indonesia begitu pesat dengan jumlah pengguna yang terus meningkat. Akan tetapi hal tersebut kurang diimbangi dengan kesadaran tentang privasi dalam kaitannya dengan big data yang dihasilkan oleh penyedia layanan. Penyedia layanan memberikan kebijakan berupa syarat dan ketentuan akan tetapi masyarakat umumnya masih rendah dalam hal memiliki kesadaran tentang privasi data pribadi mereka. Penelitian ini bertujuan untuk mengetahui solusi dari permasalahan privasi big data dalam media sosial dan dianalisis dengan teori privasi komunikasi. Metode yang digunakan dalam penelitian ini adalah metode meta-analisis yang mengolah hasil temuan dari penelitian sebelumnya. Hasil dari penelitian ini berupa solusi bagi perlindungan privasi data individu saat pembuatan, penyimpanan, dan pemrosesan data.

Kata Kunci: data besar, Indonesia, kebijakan, media sosial, privasi

Abstract

The development of social media in Indonesia is high increasing. However, this is not accompanied by awareness of privacy in its commitment to big data generated by service providers. The service provider provides an agreed policy, will provide the public about their data privacy issues. This article used Communication Privacy Management to finding solution about big data privacy problems. The method used in this study is a meta-analysis method that processes the findings from previous studies. The results of this study contain solutions for privacy protection when creating data, data storage, and processing data.

Keywords: big data, Indonesia, policy, social media, privacy

1. Pendahuluan

Di era sekarang ini, bagi yang memiliki *big data* maka akan memiliki kemampuan untuk menguasai konsumen. *Big data* dapat digunakan secara efektif untuk lebih memahami dunia dan berinovasi dalam berbagai aspek, jumlah data yang meledak telah meningkatkan potensi pelanggaran privasi. Misalnya, jejaring sosial seperti *Facebook* menyimpan semua informasi tentang kehidupan pribadi dan hubungan sosial seseorang. Situs *video sharing* seperti *YouTube* merekomendasikan tayangan kepada pengguna video berdasarkan riwayat pencarian pengguna. Kekuatan yang didorong oleh *big data* dengan mengumpulkan, menyimpan, dan menggunakan kembali informasi pribadi untuk tujuan mendapatkan keuntungan komersial, telah mengancam privasi dan keamanan. Privasi pengguna dapat dilanggar dalam kondisi berikut:

- ➔ Informasi pribadi ketika dikombinasikan dengan data eksternal dapat mengarah pada kesimpulan fakta baru tentang pengguna. Fakta-fakta itu mungkin bersifat rahasia dan tidak seharusnya diungkapkan kepada orang lain.
- ➔ Informasi pribadi terkadang dikumpulkan dan digunakan untuk menambah nilai bagi bisnis tertentu. Misalnya, kebiasaan belanja individu dapat mengungkapkan banyak informasi pribadi.
- ➔ Data sensitif disimpan dan diproses di lokasi yang tidak diamankan dengan baik dan kebocoran data dapat terjadi selama fase penyimpanan dan pemrosesan.

Masyarakat mengungkapkan informasi pribadi secara sadar atau tidak sadar, mau atau tidak mau, ketika mereka melakukan kegiatan sehari-hari misalnya berbelanja bahan makanan, berkomunikasi dengan anggota keluarga, membayar pajak, membaca berita, mendengarkan musik, membaca *e-book*, berbagi foto, dan sebagainya. Banyak orang memilih untuk mengungkapkan informasi tentang kehidupan pribadi mereka di situs jejaring sosial. *Great privacy giveaway* (Allen, 2013, 847) atau "eksibisionisme media" (Nissenbaum 2010,

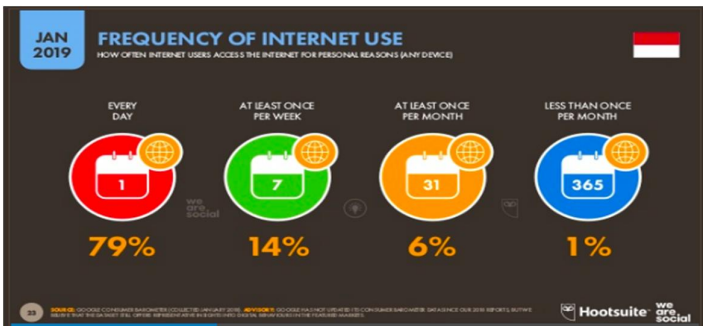
106), tidak ada perbedaan dengan alasan untuk mengungkapkan informasi tetapi tercatat bahwa hampir tidak mungkin untuk melakukan sebagian besar kegiatan sehari-hari tanpa mengungkapkan informasi pribadi dan menyediakan makanan bagi "pialang" data dan organisasi *big data*. Informasi pribadi sering diungkapkan secara sukarela, karena orang bersedia memberikan informasi pribadi saat berinteraksi dengan entitas digital, dalam beberapa kasus mereka bahkan mungkin secara eksplisit telah menyetujui pihak yang mengumpulkan informasi pribadi mereka. Hal itu cukup beralasan bahwa seseorang telah melepaskan hak mereka mengenai privasi sehubungan dengan informasi yang disebarkan. Orang-orang telah dengan sukarela memberikan informasi pribadi mereka dalam kegiatan sehari-hari mereka.

Solove (2013) menyatakan bahwa inti dari masalah privasi adalah dilema persetujuan. Banyak praktik dan teori privasi bergantung pada gagasan bahwa orang memiliki hak atau peluang untuk menyetujui untuk memberikan data pribadi yang diminta. Solove (2013) menunjukkan dalam analisisnya bahwa pendekatan dasar untuk melindungi privasi sebagian besar tetap tidak berubah sejak tahun 1970-an dan bahwa pendekatan persetujuan terhadap privasi sudah tidak lagi berarti dalam masyarakat informasi jaringan saat ini. Dalam praktiknya, orang setuju untuk memberikan informasi pribadi tanpa banyak berpikir, tanpa membaca atau sepenuhnya memahami formulir persetujuan, dan lebih jauh lagi, formulir persetujuan seringkali agak panjang dan ditulis dalam bahasa legalistik yang hanya sedikit orang mengerti.

Di sisi lain, teknologi digital berkembang sedemikian pesat hingga muncul berbagai macam aplikasi yang dibutuhkan oleh user. Berikut merupakan data penggunaan internet secara global.

Gambar 1. Konsumsi Digital di Dunia (sumber: https://wearesocial.com/blog/2019/01/digital-2019-global-internet-use-accelerates)

Berikut ini merupakan data yang diperoleh dari situs *wearesocial.com* mengenai penggunaan teknologi digital di Indonesia.



Gambar 2. Frekuensi Penggunaan Internet di Indonesia (sumber: https://datareportal.com/reports/digital-2019-indonesia)

Di Indonesia pengguna internet aktif sebagian besar menggunakan internet setiap hari sebanyak 79%. Hal ini menandakan bahwa pengguna internet di Indonesia merupakan pengguna aktif yang biasanya dapat menjadi sasaran oleh pemasar dalam memasarkan produknya. Saat ini berbagai aplikasi berkembang begitu cepat yang membuat kondisi dunia digital di Indonesia kian ramai.

Perkembangan dunia digital tidak lepas dari adanya media sosial. Hampir setiap hari sebagian masyarakat memiliki ketergantungan hidup dengan media sosial. Melihat kenyataan tersebut, dapat dilihat bahwa data yang ada begitu

banyak dan dapat diolah atau dimanfaatkan. Media sosial yang memiliki data yang besar ini memiliki perubahan yang begitu cepat dan dimana rentang waktu perubahannya memiliki manfaat yang relevan dalam berbagai bidang jika kemudian data tersebut digunakan. Jika ditelaah lebih jauh, media sosial memiliki sumber *big data* yang bernilai baik di bidang ekonomi, politik, budaya dan lain sebagainya yang digunakan sebagai prediksi maupun sumber data mengenai potensi tertentu.

Di era internet saat ini, pihak yang memiliki wewenang dan akses data dapat mengumpulkan data yang dibutuhkan dalam skala besar, tetapi terdapat masalah yang harus dihadapi yaitu ekstraksi pengetahuan yang dihasilkan oleh data tertentu dapat menyebabkan masalah pada privasi pengguna. Pengumpulan dan analisis data yang dilakukan dalam media sosial memiliki banyak tujuan, misalnya dalam hal yang berkaitan dengan pelanggan, data tertentu dapat digunakan untuk meningkatkan rekomendasi terhadap produk tertentu secara personal. Terkait dengan masalah privasi, data yang dikumpulkan media sosial dapat berupa informasi pribadi yang sensitif bila dilihat dari sumber datanya.

Ada berbagai jenis media sosial, salah satu diantaranya adalah *Facebook.* Dalam penelitian yang telah dilakukan oleh Jennifer Golbeck dan Matthew Mauriello menunjukkan bahwa rata-rata pengguna salah satu jenis media sosial *Facebook* secara signifikan telah mengesampingkan jumlah data yang mereka izinkan untuk diakses oleh aplikasi pihak ketiga. Selain itu mereka juga menemukan bahwa pengguna cenderung mengabaikan mengenai ketentuan privasi dan kebijakan di *web* yang diakses.

Media sosial sendiri lebih rentan terhadap pelanggaran privasi sehingga saat ini peneliti maupun profesional bidang teknologi informasi akan memperhatikan teknologi, model bisnis, dan upaya regulasi sehingga tujuan penyedia layanan maupun pengguna dapat terpenuhi.

Selain itu, 85 persen pengguna Internet ingin memahami lebih banyak tentang pengumpulan data, dan 88 persen responden ingin mengontrol data

yang dikumpulkan dari perangkat mereka. Akhirnya, survei tersebut mengungkapkan bahwa 87 persen pengguna internet khawatir tentang jenis informasi pribadi yang dikumpulkan (Perera, et. al., 2015).

Saat ini, konsumen menyadari bahwa ketika mereka menggunakan layanan online gratis (email, jaringan sosial, dan *feed* berita, misalnya), mereka secara otomatis menjadi sumber data untuk kegiatan bisnis, yaitu dapat menganalisis data untuk meningkatkan kepuasan pelanggan. Hal yang lebih buruk lagi, data dapat digunakan pihak ketiga untuk analisis lebih mendalam misalnya dalam kasus bocornya 235 juta pengguna *Youtube*, Instagram dan TikTok dimana kebocoran data tersebut berasal dari Deep Social. Data yang terungkap atau bocor adalah nomor telepon dan alamat email. Kebocoran tersebut terjadi dan dilaporkan pada tanggal 1 Agustus 2020 oleh Bob Diachenko (Franedya, 2020). Dilihat dari kejadian tersebut, hingga saat ini masih terjadi pengambilan data pribadi secara illegal sehingga hal tersebut dapat mengganggu privasi atau keamanan data pribadi pelanggan yang tidak diinginkan. Berdasarkan hal tersebut, penulis ingin mengetahui lebih jauh apa solusi dari permasalahan perlindungan privasi data dalam media sosial.

2. Tinjauan Pustaka

2.1. Teori Manajemen Privasi Komunikasi

Communication Privacy Management menjelaskan tentang proses-proses negosiasi mengenai pengungkapan informasi privat (Petronio, 2002: 3). Setiap orang memiliki pilihan untuk mengungkap atau menutupi informasi yang mereka miliki berdasarkan kondisi yang mereka anggap penting, dan mereka mempunyai hak untuk memiliki dan mengatur akses ke informasi pribadimereka (Petronio, 2002: 2). Terdapat 6 prinsip utama CPM yaitu *Public-Private Dialectical Tension, Private-Information, Privacy Rules, Boundaries, Boundary Coordination, Boundary Turbulence*

Lebih lanjut CPM dapat diaplikasikan dalam konteks personal, kelompok, organisasi serta dalam konteks komunikasi digital.

2.2 Privasi

Tavani (2008) mengemukakan bahwa ada empat jenis privasi, yang masing-masing memiliki fokus unik. Gagasan privasi Warren dan Brandeis dapat disebut privasi fisik (Tavani 2008, 135), yang merupakan kebebasan dari intrusi (fisik). Pada poin kedua dan ketiga, masing-masing adalah privasi keputusan, yang merupakan kebebasan dari gangguan yang mempengaruhi keputusan penting, dan privasi psikologis, yang berkaitan dengan perlindungan pikiran intim seseorang. Pada poin keempat, privasi informasi, adalah yang dibahas khusus dalam tulisan ini. Jenis privasi ini sering dikaitkan dengan definisi Westin (1967) dari tahun 1960-an: Privasi adalah klaim individu, kelompok, atau lembaga untuk menentukan sendiri kapan, bagaimana, dan sejauh mana informasi tentang mereka dikomunikasikan kepada orang lain. Seperti disebutkan sebelumnya, dua pendekatan umum untuk mengkonseptualisasikan privasi informasi adalah akses yang terbatas (Tavani, 2008) dan kontrol. Dalam istilah akses yang terbatas, seseorang dapat membatasi orang lain dari informasi tentang dirinya sendiri. Gagasan dasar di sini adalah bahwa seorang individu dapat mengatur zona atau konteks pribadi dimana informasi pribadi diadakan dan individu tersebut dapat menikmati privasi ketika ia dapat membatasi atau membatasi orang lain dari mengakses informasi pribadi yang diadakan di zona-zona tersebut. Kontrol terkait dengan akses terbatas yang berkaitan dengan kemampuan individu untuk mengendalikan siapa yang memiliki akses ke informasi tentang individu. Elemen utama dalam kontrol adalah peran yang dimainkan oleh pilihan individu dan kemampuan individu untuk mengontrol apakah akses ke informasi tentang diri mereka diberikan atau dibatasi kepada orang lain. Model privasi data menggeser fokus dari pengumpulan data ke

pemrosesan data dan analisis. Pengumpulan data berorientasi ontologis; dimana hal ini berfokus pada data yang mewakili fakta tentang keadaan di dunia: orang dan kegiatan dan keterkaitan antara tempat, waktu, orang lain, kegiatan dan tujuannya. Pemrosesan dan analisis data secara epistemologis berorientasi dan berfokus pada fakta atau kenyataan bahwa data dapat dihasilkan saat mereka diproses dan dianalisis.

Tiga model privasi yaitu pengawasan, penangkapan, dan datafikasi memiliki makna yang saling melengkapi karena model yang berbeda yang membantu kita untuk memahami dan menghargai gagasan privasi, masing-masing model memiliki sudut pandang yang berbeda dan berfokus pada fitur spesifik dari fenomena sosioteknologi yang sedang diselidiki. Model pengawasan berfokus pada ketegangan antara pengamat dan yang diawasi, antara ruang publik dan pribadi, dan pada hubungan kekuasaan saat ini. Model penangkapan berfokus pada kodifikasi kegiatan, sifat sosioteknik teknologi komputer, dan pada tujuan pengumpulan data yang tidak jelas. Model datafikasi berfokus pada pembuatan informasi pribadi anonim yang baru, reinterpretasi dan analisis statistik data, dan sifat informasi pribadi yang dikomodifikasi. Secara bersama-sama, ketiga model ini memberikan pendekatan holistik yang kuat untuk privasi dalam masyarakat informasi jaringan; secara individual, mereka menekankan aspek yang berbeda dan menyoroti fitur yang berbeda.

2.3 Media Sosial

Russo, Watkins, Kelly, dan Chan (2008) mendefinisikan media sosial sebagai pihak yang memfasilitasi komunikasi *online*, jaringan, dan atau kolaborasi. Kaplan dan Haenlein (2010) memberikan definisi singkat yang sama tentang media sosial sebagai "sekelompok aplikasi berbasis Internet yang dibangun di atas fondasi ideologis dan teknologi Web 2.0, dan yang memungkinkan pembuatan dan pertukaran konten. Lewis

(2010) mendefinisikan media sosial hanya berfungsi sebagai label untuk teknologi digital yang memungkinkan orang untuk terhubung, berinteraksi, memproduksi dan berbagi konten.

Howard dan Parks (2012) memberikan definisi yang lebih kompleks dari media sosial yang terdiri dari tiga bagian: (a) infrastruktur informasi dan alat yang digunakan untuk memproduksi dan mendistribusikan konten; (b) konten yang mengambil bentuk digital dari pesan pribadi, berita, ide, dan produk budaya; dan (c) orang, organisasi, dan industri yang memproduksi dan menggunakan konten digital.

Mereka lebih lanjut menetapkan bahwa media sosial sering dilambangkan dalam literatur, bukan oleh sifat dan karakteristik mereka tetapi dengan hanya menggunakan aplikasi spesifik seperti *Facebook* atau *YouTube.*

Dalam hubungan masyarakat, Kent (2010) secara luas mendefinisikan media sosial sebagai saluran komunikasi interaktif apa pun yang memungkinkan interaksi dan umpan balik dua arah, lebih lanjut menentukan media sosial modern dicirikan oleh potensi mereka untuk interaksi waktu nyata, mengurangi anonimitas, rasa kedekatan, waktu respons yang singkat, dan kemampuan untuk 'menggeser waktu,' atau terlibat dalam jejaring sosial kapan saja sesuai dengan masing-masing anggota.

Media sosial sering dikonseptualisasikan secara tekno-sentris, berdasarkan pada perangkat tertentu, sering dianggap identik dengan Web 2.0 atau web kolaboratif (mis., Agichtein, Castillo, Donato, Gionis, & Mishne, 2008). Web 2.0 merujuk pada alat kolaborasi berbasis web yang mengandalkan konten yang dibuat pengguna yang terus berkembang dan meningkat (O'Reilly, 2005). Hal yang lebih problematis adalah penggabungan "media sosial" dan "situs jejaring sosial." Boyd dan Ellison (2007) mendefinisikan situs jejaring sosial (SNS) sebagai layanan berbasis

web yang memungkinkan individu untuk (1) membangun publik atau semi profil publik dalam sistem terikat, (2) mengartikulasikan daftar pengguna lain dengan siapa mereka berbagi koneksi, dan (3) melihat dan melintasi daftar koneksi mereka dan yang dibuat oleh orang lain dalam sistem. Definisi ini sering secara salah diterapkan sebagai definisi media sosial yang menyeluruh meskipun SNS menurut sifatnya adalah alat media sosial, tidak semua media sosial pada dasarnya adalah SNS.

Kurangnya definisi yang stabil, juga menimbulkan masalah yang signifikan untuk mengetahui media sosial di masa depan. Tanpa melihat secara objektif tentang apa itu media sosial, akan sulit untuk memahami bagaimana mendekati dan berteori masalah yang terjadi dalam media sosial dari perspektif komunikasi dan sebagainya.

Definisi media sosial yang baru adalah saluran berbasis Internet yang memungkinkan pengguna untuk berinteraksi secara mandiri menampilkan diri, baik secara *real-time* atau tidak serempak, dengan audiens yang luas dan sempit sehingga memperoleh nilai dari konten buatan pengguna dan persepsi interaksi dengan orang lain.

Media sosial beroperasi melalui internet yang lebih luas. Internet merujuk ke jaringan komputer yang saling terhubung di seluruh dunia dan merujuk terutama untuk infrastruktur sistem, sedangkan *World Wide Web* adalah salah satu dari banyak aplikasi yang menggunakan infrastruktur Internet untuk berkomunikasi melalui *hyperlink audiovisual* dan diakses melalui *browser*. Semakin banyak pengembang yang beralih dari alat web berbasis *browser* untuk memasukkan aplikasi mandiri yang tidak mengharuskan web berfungsi.

Memecah definisi media sosial dari gagasan terkini tentang alat Web 2.0 seperti *Facebook* dan Instagram memungkinkan untuk dimasukkannya alat yang melampaui pengertian saat ini tentang *web* dan aplikasi *online* namun masih termasuk alat sosial yang beroperasi pada

intranet pribadi *multisite* organisasi (yang terhubung melalui Internet) seperti IBM Beehive (Thom-Santelli, Millen, & DiMicco, 2010). Pengembang media sosial terus merangkul aplikasi dengan menggunakan web sebagai platform, media sosial dapat mengandalkan aplikasi internet lainnya, termasuk protokol *transfer file* dan *streaming media* untuk memfasilitasi komunikasi yang terjadi.

2.4 Big Data

Big data sangat identik dengan 3V yaitu *volume, velocity, variety* (Laney, 2001). *Volume* merupakan ukuran data yang sangat besar. *Velocity* atau kecepatan terkait dengan kecepatan dimana data dibuat dan tersedia dan yang terakhir yaitu *variety* mengacu pada format yang berbeda. Singkatnya, *big data* merupakan data yang berukuran sangat besar yang masuk dengan ukuran dan format yang berbeda-beda.

Siklus hidup *big data* terdiri dari 3 bagian yaitu;

a. Pembuatan data: Data dapat dihasilkan dari berbagai sumber yang didistribusikan. Jumlah data yang dihasilkan oleh manusia dan mesin telah meledak dalam beberapa tahun terakhir. Misalnya, setiap hari 2,5 *quintillion byte* data dihasilkan di web dan 90 persen dari seluruh data di dunia dihasilkan dalam beberapa tahun terakhir. *Facebook*, situs jejaring sosial sendiri menghasilkan 25TB data baru setiap hari. Biasanya, data yang dihasilkan besar, beragam, dan kompleks sehingga sulit bagi sistem tradisional untuk menanganinya. Data yang dihasilkan biasanya dikaitkan dengan domain tertentu seperti bisnis, internet, penelitian, dan lain-lain.
b. Penyimpanan data: Fase ini mengacu pada menyimpan dan mengelola set data skala besar. Sistem penyimpanan data terdiri dari dua bagian yaitu, infrastruktur perangkat keras dan manajemen data (Hu, et., al. 2006). Infrastruktur perangkat keras mengacu pada pemanfaatan sumber daya teknologi informasi dan komunikasi (TIK) untuk berbagai tugas. Manajemen data mengacu pada set perangkat lunak yang digunakan di

atas infrastruktur perangkat keras untuk mengelola dan meminta set data skala dalam skala besar.

c. Pemrosesan data: Fase pemrosesan data pada dasarnya mengacu pada proses pengumpulan data, pengiriman data, pra-pemrosesan, dan penggalian informasi yang berguna. Pengumpulan data diperlukan karena data mungkin berasal dari berbagai sumber yang berbeda, yaitu situs yang berisi teks, gambar, dan video. Dalam fase pengumpulan data, data diperoleh dari lingkungan produksi data spesifik menggunakan teknologi pengumpulan data khusus. Dalam data fase transmisi, setelah mengumpulkan data mentah dari lingkungan produksi data tertentu kita memerlukan mekanisme transmisi kecepatan tinggi untuk mengirimkan data ke penyimpanan yang tepat untuk berbagai jenis aplikasi analitik. Akhirnya, fase pra-pemrosesan bertujuan untuk menghilangkan bagian data yang tidak berarti dan berlebihan sehingga lebih banyak ruang penyimpanan dapat dihemat. Metode analitik spesifik data dan domain berlebihan digunakan oleh banyak aplikasi untuk mendapatkan informasi yang bermakna. Meskipun bidang yang berbeda dalam analisis data memerlukan karakteristik data yang berbeda, beberapa bidang ini dapat memanfaatkan teknologi mendasar yang serupa untuk memeriksa, mengubah, dan memodelkan data untuk mengekstraksi nilai darinya. Penelitian analitik data yang muncul dapat diklasifikasikan ke dalam enam bidang teknis berikut: analisis data terstruktur, analisis teks, analisis multimedia, analisis web, analisis jaringan, dan analisis seluler.

3. Metode Penelitian

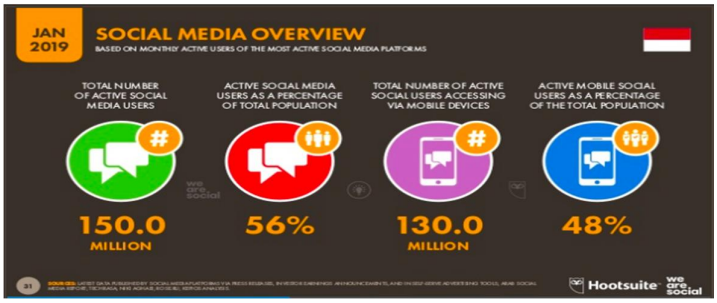
Penelitian ini dilakukan dengan menganalisis sebuah analisis yang sudah ada sebelumnya dimana penelitian terdahulu yang berkaitan dengan proteksi privasi media sosial yang secara langsung berkaitan dengan *big data*. Penelitian ini selanjutnya berfokus pada data-data yang dikumpulkan dari penelitian lain. Penelitian ini menggunakan konsep meta analisis yang

merupakan analisis integratif sekunder. Menurut Glass (1981), analisis sekunder merupakan analisis yang dilakukan kembali terhadap suatu data yang sudah ada dengan tujuan untuk menjawab pertanyaan penelitian dengan data yang sudah ada sebelumnya. Studi mengenai proteksi privasi media sosial ini dilakukan dengan cara menganalisis data yang dari penelitian yang dilakukan sebelumnya.

Hasil analisis penelitian sebelumnya mengenai proteksi privasi media sosial dipakai sebagai dasar untuk menerima atau mendukung hipotesis, menolak hipotesis yang diajukan oleh beberapa peneliti (Sugiyanto, 2004). Hal ini dilakukan karena untuk mengkaji keajegan atau ketidakjegan hasil dari sebuah penelitian yang disebabkan semakin sering terjadinya replikasi atau verifikasi penelitian, yang seringnya memperbesar adanya variasi hasil penelitian.

4. Hasil dan Pembahasan

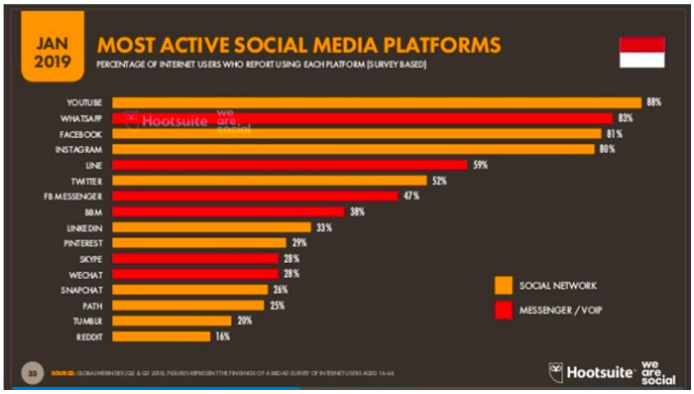
Berikut merupakan hasil pengguna sosial media di Indonesia;



Gambar 3. Pengguna Media Sosial (Sumber: https://datareportal.com/reports/digital-2019-indonesia)

Berdasarkan data diatas, kita dapat melihat bahwa pengguna media di Indonesia cukup tinggi dengan pengguna sosial media yang aktif sebanyak 56%. Hal ini dapat disimpulkan bahwa setengah lebih populasi masyarakat Indonesia menggunakan internet secara aktif. Berdasarkan data diatas juga dapat dilihat bahwa masyarakat sebagian besar mengakses internet dari perangkat *mobile* mereka, sehingga hal ini dapat disimpulkan bahwa pengguna perangkat *mobile* di

Indonesia dinilai cukup tinggi. Pengguna media sosial di perangkat mobile yang setiap saat selalu aktif sebesar 48% dari total pengguna media sosial. Hal tersebut dapat dikategorikan bahwa setengah dari populasi pengguna media sosial *mobile* memiliki ketergantungan hidup terhadap media sosial.



Gambar 4. *Platform* Media Sosial (sumber: https://datareportal.com/reports/digital-2019-indonesia)

Data terbaru yang didapatkan pada tahun 2019 menunjukkan bahwa pengguna *Youtube* di Indonesia adalah yang paling tinggi jika dibandingkan dengan *social network site* (SNS) yang lain yang kemudian diikuti oleh *whatsapp.* Hal tersebut menunjukkan bahwa frekuensi orang indonesia menonton video tergolong cukup tinggi dan yang kedua mereka berhubungan melalui pesan yang juga tergolong sering. Media sosial yang sering dipakai yang ketiga adalah *facebook* yang dilanjutkan dengan instagram. Instagram merupakan *platform* yang menfasilitasi pengguna berbagi dalam bentuk gambar. Selanjutnya diikuti *twitter* dan *linkedin.* Tren pengguna media sosial di Indonesia cenderung meningkat. Hal tersebut seharusnya menjadi pemicu untuk meningkatkan privasi pengguna media sosial yang terkait dengan keamanan *big data*.

Big data membutuhkan komputasi dan penyimpanan yang besar, yang mana hal tersebut membawa kebutuhan untuk *cloud computing*. *Cloud computing* mendorong perusahaan untuk mengadopsi *cloud,* karena banyak keuntungan yang ditawarkan, seperti penghematan biaya dan juga menawarkan kekuatan pemrosesan dan kemampuan penyimpanan dalam skala besar.

Teknologi yang digunakan dalam *cloud computing* seperti virtualisasi, penyimpanan dan pemrosesan memungkinkan untuk melakukan tugas yang sulit jika dilakukan secara konvensional. Umumnya beberapa pihak ragu mentransfer data pribadi mereka atau data sensitif ke *cloud* kecuali mereka yakin jika hal itu data mereka akan aman di *cloud*. ada beberapa tantangan untuk membangun *big data* yang dapat dipercaya dan aman sistem penyimpanan dan pemrosesan di *cloud* yang dapat diikuti (Xiao, 2013).

1. Pengalihdayaan: Untuk mengurangi modal dan operasional, organisasi saat ini lebih memilih untuk melakukan *outsourcing* data mereka ke cloud. Namun, *outsourcing* data ke *cloud* juga berarti bahwa pelanggan akan kehilangan kontrol fisik pada data mereka. Hilangnya kontrol atas data telah menjadi salah satu penyebab utama kerawanan *cloud*. Ketidakamanan ini dapat menyebabkan kerusakan serius pada privasi pelanggan *cloud computing*. Masalah-masalah ini bisa saja diatasi dengan menyediakan lingkungan komputasi yang aman dan penyimpanan data. Selain itu, data *outsourcing* juga harus diverifikasi kepada pelanggan dalam hal kerahasiaan dan integritas.
2. Multi-tenancy: Virtualisasi memungkinkan berbagi *platform cloud* yang sama oleh banyak pelanggan. Data milik pengguna *cloud* yang berbeda mungkin ditempatkan pada penyimpanan fisik yang sama oleh beberapa sumber kebijakan alokasi. Dalam lingkungan seperti itu, itu relatif mudah bagi pengguna jahat untuk secara ilegal mengakses data yang bukan miliknya. Serangkaian masalah dapat terjadi dalam lingkungan seperti itu, seperti pelanggaran data dan pelanggaran perhitungan sehingga sangat penting untuk mekanisme desain untuk menangani potensi privasi dan risiko keamanan.
3. Komputasi masif: Karena kemampuan *cloud* komputasi untuk menangani penyimpanan big data dan perhitungan intens,

mekanisme tradisional untuk melindungi privasi individu tidak memadai.

4.1 Perlindungan Privasi Bagian Pembuatan Data

Privasi dalam pembuatan data dapat diklasifikasikan ke dalam pembuatan data aktif dan pembuatan data pasif. Pembuatan data aktif berarti bahwa pemilik data bersedia memberikan data kepada pihak ketiga, sementara pembuatan data pasif mengacu pada situasi di mana data dihasilkan oleh aktivitas online pemilik data misalnya data penelusuran dimana pemilik data mungkin tidak menyadari bahwa data sedang dikumpulkan oleh pihak ketiga. Tantangan utama bagi pemilik data adalah bagaimana ia dapat melindungi datanya dari pihak ketiga mana pun yang mungkin ingin mengumpulkannya. Pemilik data ingin menyembunyikan informasi pribadi atau yang bersifat sensitif sebanyak mungkin dan khawatir tentang seberapa besar kontrol yang dapat ia miliki atas informasi tersebut. Hal ini dapat meminimalkan risiko pelanggaran privasi selama pembuatan data dengan membatasi akses atau memalsukan data (Xu, 2014).

Untuk memproteksi data dari pihak ketiga, dapat dilakukan dengan cara pembatasan akses. Jika pemilik data berpikir bahwa data tersebut dapat mengungkapkan informasi sensitif yang tidak seharusnya dibagikan, ia dapat menolak untuk memberikan data tersebut. Untuk itu, pemilik data harus mengadopsi metode kontrol akses yang efektif sehingga data dapat dicegah agar tidak dicuri oleh pihak ketiga. Jika pemilik data menyediakan data secara pasif, beberapa tindakan dapat diambil untuk memastikan privasi, seperti ekstensi anti-pelacakan, pemblokir iklan dan alat enkripsi (Xu, 2014). Penggunakan alat ini memudahkan seseorang untuk dapat secara efektif membatasi akses ke data yang bersifat sensitif. Untuk kemudahan penggunaan, sebagian besar alat ini dirancang sebagai ekstensi peramban.

Selain itu, ada beberapa cara alternatif, seperti menggunakan perangkat lunak *anti-malware* dan anti-virus untuk melindungi data yang disimpan secara

digital di komputer atau laptop. Alat-alat ini dapat membantu melindungi data pribadi pengguna dengan membatasi akses meskipun tidak ada jaminan bahwa data yang bersifat sensitif milik seseorang sepenuhnya dilindungi dari sumber yang tidak dapat dipercaya.

Terkadang dalam lain keadaan, tidak mungkin untuk mencegah akses data yang bersifat sensitif. Dalam hal ini, data dapat terdistorsi menggunakan alat tertentu sebelum data diambil oleh beberapa pihak ketiga. Jika data terdistorsi, informasi sebenarnya tidak dapat dengan mudah diungkapkan. Teknik-teknik berikut digunakan oleh pemilik data untuk memalsukan data diantaranya (Xu, 2014).

5. Alat *Socketpuppet* digunakan untuk menyembunyikan identitas *online* individu dengan penipuan. Aktivitas online individu yang sebenarnya dirahasiakan dengan menciptakan identitas palsu dan berpura-pura menjadi orang lain. Dengan menggunakan beberapa *Socketpuppets*, data milik satu individu tertentu akan dianggap milik individu yang berbeda. Cara tersebut menyebabkan para pengumpul data tidak akan memiliki cukup pengetahuan untuk mengaitkan berbagai soket soket yang berbeda dengan satu orang. Oleh karena itu, aktivitas pengguna yang sebenarnya tidak diketahui orang lain dan informasi pribadi tidak dapat ditemukan dengan mudah.
6. Alat keamanan tertentu dapat digunakan untuk menutupi identitas individu, seperti MaskMe. Ini memungkinkan pengguna untuk membuat alias informasi pribadi mereka seperti alamat email atau nomor kartu kredit. Pemilik data dapat menggunakan topeng ini setiap kali informasi dibutuhkan. Ini sangat berguna ketika pemilik data perlu memberikan rincian kartu kredit selama belanja online.

B. Perlindungan Privasi Bagian Penyimpanan Data

Menyimpan data yang "*high volume*" bukanlah tantangan besar karena adanya kemajuan teknologi penyimpanan data seperti *booming* dalam *cloud computing*. Namun, mengamankan data menjadi tantangan tersendiri. Jika sistem penyimpanan *big data* terganggu, bisa jadi sangat berbahaya karena informasi pribadi individu dapat diungkapkan. Oleh karena itu, perlu dipastikan bahwa data yang disimpan dilindungi dari ancaman tersebut. Dalam sistem informasi modern, pusat data memainkan peran penting dalam melakukan pengambilan sejumlah besar data. Misalnya, suatu aplikasi mungkin memerlukan beberapa set data dari pusat data yang berbeda dan karena itu menghadapi tantangan perlindungan privasi.

Mekanisme keamanan konvensional untuk melindungi data dapat dibagi menjadi empat kategori. Diantaranya adalah skema keamanan data tingkat *file*, skema keamanan data tingkat *database*, skema keamanan tingkat media dan skema enkripsi tingkat aplikasi (Hongbing, 2015). Mekanisme konvensional untuk melindungi keamanan data (Cao, 2014) dan privasi (Soundararajan, 2014), untuk arsitektur penyimpanan penyimpanan yang ada (yaitu, penyimpanan terpasang langsung, penyimpanan jaringan terpasang dan jaringan area penyimpanan) telah menjadi area penelitian yang sangat panas tetapi mungkin tidak dapat langsung diterapkan ke platform analitik big data (Troppens, 2011). Menanggapi sifat analitik big data 3V, infrastruktur penyimpanan harus *scalable*. Idealnya, memiliki kemampuan untuk dikonfigurasi secara dinamis untuk mengakomodasi beragam aplikasi. Salah satu teknologi yang menjanjikan untuk memenuhi persyaratan ini adalah virtualisasi penyimpanan, diaktifkan oleh paradigma *cloud computing* yang muncul (Mell, 2013). Virtualisasi penyimpanan adalah proses di mana beberapa perangkat penyimpanan jaringan digabungkan menjadi apa yang tampak sebagai perangkat penyimpanan tunggal. Namun, menggunakan layanan *cloud* yang ditawarkan oleh penyedia cloud berarti bahwa data organisasi akan diserahkan kepada pihak ketiga seperti penyedia cloud. Ini dapat memengaruhi

privasi data. Oleh karena itu, dalam tulisan ini diskusi dibatasi untuk privasi data saat disimpan di *cloud.*

Pendekatan yang dapat dilakukan dalam preservasi privasi misalnya penyimpanan di cloud. Ketika data disimpan di cloud, keamanan data terutama memiliki tiga dimensi, kerahasiaan, integritas, dan ketersediaan (Xiao, 2013). Dua yang pertama terkait langsung dengan privasi data yaitu, jika kerahasiaan atau integritas data dilanggar, hal tersebut akan berdampak langsung pada privasi pengguna.

Persyaratan dasar untuk sistem penyimpanan *big data* adalah untuk melindungi privasi individu. Ada beberapa mekanisme yang ada untuk memenuhi persyaratan itu. Misalnya, pengirim dapat mengenkripsi datanya menggunakan *public key encryption* (PKE) sedemikian rupa sehingga hanya penerima yang valid yang dapat mendekripsi data. Pendekatan untuk menjaga privasi pengguna saat data disimpan di *cloud* adalah sebagai berikut.

1. ABE (*Attribute Based Encryption*) adalah teknik enkripsi yang memastikan privasi big data ujung ke ujung dalam sistem penyimpanan cloud. Dalam kebijakan akses ABE ditentukan oleh pemilik data dan data dienkripsi berdasarkan kebijakan tersebut. Data hanya dapat didekripsi oleh pengguna yang atributnya memenuhi kebijakan akses yang ditentukan oleh pemilik data. Ketika berhadapan dengan *big data*, seseorang mungkin sering perlu mengubah kebijakan akses data karena pemilik data mungkin harus membaginya dengan organisasi yang berbeda. Skema kontrol akses berbasis atribut saat ini (Yang, et. al., 2013), tidak mempertimbangkan pembaruan kebijakan. Pembaruan kebijakan adalah tugas yang sangat menantang dalam sistem kontrol akses berbasis atribut. Alasan untuk itu adalah setelah data di-*outsourcing*-kan ke *cloud*, pemilik data tidak akan menyimpan salinan lokal dalam sistem. Jika pemilik data ingin memperbarui kebijakan, ia harus mentransfer data kembali ke sistem

lokal, mengenkripsi ulang data di bawah kebijakan baru dan menyimpannya kembali di *server cloud*. Proses ini memiliki *overhead* komunikasi yang sangat tinggi dan biaya komputasi yang tinggi. Untuk mengatasi masalah pembaruan kebijakan, baru-baru ini Yang mengusulkan kebijakan pembaruan metode *outsourcing* yang aman dan dapat diverifikasi. Dalam, pemilik data tidak perlu mengambil semua data dan mengenkripsi ulang. Sebaliknya pemilik data dapat mengirim kueri ke *cloud* untuk memperbarui kebijakan, dan *server cloud* dapat memperbarui kebijakan secara langsung tanpa mendekripsi data.

2. IBE (*Identity Based Encryption*) adalah alternatif yang dapat digunakan untuk menyederhanakan manajemen kunci dalam infrastruktur kunci publik (PKI) berbasis sertifikat dengan menggunakan identitas manusia seperti alamat email atau alamat IP sebagai kunci publik. Untuk menjaga anonimitas pengirim dan penerima, skema IBE diusulkan. Penggunaan metode ini menyebabkan sumber dan tujuan data dapat dilindungi secara pribadi. Skema enkripsi seperti IBE dan ABE tidak mendukung pembaruan penerima *ciphertext.* Ada beberapa pendekatan untuk memperbarui penerima *ciphertext*. Misalnya, pemilik data dapat menggunakan mode dekripsi lalu enkripsi ulang namun jika *big data* karena sebagian besar kasus ketika berhadapan dengan *big data*, dekripsi dan re-enkripsi bisa sangat memakan waktu dan mahal karena perhitungan *overhead*. Dalam mode ini, pemilik data harus *online* setiap saat. Pendekatan lain untuk memperbarui penerima *ciphertext* adalah mendelegasikan tugas ini kepada pihak ketiga yang tepercaya dengan pengetahuan tentang kunci dekripsi pemilik data. Pendekatan ini memiliki beberapa kelemahan seperti skema bergantung pada kepercayaan penuh dari pihak ketiga dan juga anonimitas penerima *ciphertext* tidak dapat dicapai karena pihak

ketiga perlu mengetahui informasi tentang tanda terima untuk melanjutkan enkripsi ulang. Enkripsi ulang *proxy* digunakan dalam pengaturan IBE. Enkripsi ulang *proxy* berbasis identitas anonim (IBPRE) diperkenalkan tetapi pekerjaan itu hanya mendukung satu kali pembaruan penerima ciphertext, sementara dalam praktiknya beberapa penerima pembaruan diinginkan. Di sisi lain, pekerjaan menyediakan mode berbagi semua atau tidak sama sekali yang membatasi fleksibilitas. Liang et al. mengusulkan skema enkripsi ulang proxy berdasarkan identitas anonim dengan properti berikut: informasi identitas pengirim dan penerima bersifat anonim dan penerima *ciphertext* dapat diperbarui beberapa kali.

3. *Homomorphic Encryption*. *Cloud storage* publik lebih rentan terhadap pelanggaran privasi karena *multi-tenancy* dan virtualisasi. Pengguna cloud dapat berbagi ruang fisik yang sama dan dalam skenario seperti itu kemungkinan kebocoran data sangat tinggi. Salah satu cara untuk melindungi data di *cloud* adalah dengan mengenkripsi data dan menyimpannya di *cloud* dan memungkinkan *cloud* melakukan perhitungan dibandingkan data yang dienkripsi. Enkripsi sepenuhnya homomorfik adalah jenis enkripsi yang memungkinkan fungsi untuk dihitung pada data yang dienkripsi. Hanya diberikan enkripsi pesan, seseorang dapat memperoleh enkripsi fungsi pesan itu dengan menghitung langsung pada enkripsi. Enkripsi homomorfik memberikan privasi penuh tetapi ia datang pada biaya kompleksitas komputasi dan kadang-kadang sangat sulit untuk diterapkan dengan teknologi yang ada.
4. *Storage Encryption*. Cheng mengusulkan skema untuk penyimpanan big data yang aman di *cloud*. Dalam skema yang diusulkan, big data pertama-tama dipisahkan menjadi banyak bagian berurutan dan kemudian setiap bagian disimpan pada media penyimpanan yang

berbeda yang dimiliki oleh penyedia penyimpanan *cloud* yang berbeda. Untuk mengakses data, bagian-bagian yang berbeda pertama-tama dikumpulkan bersama-sama dari pusat data yang berbeda dan kemudian dikembalikan ke bentuk aslinya sebelum disajikan kepada pemilik data. Dalam skema ini, big data yang disimpan di *cloud* diklasifikasikan ke dalam data publik dan data rahasia.

5. Penggunaan *Cloud Hybrid*. Menurut National Institute of Standard Technology (NIST), *cloud* dapat digunakan oleh tiga model berikut: *cloud* pribadi (dimiliki dan diakses hanya oleh perusahaan penyedia), *cloud* publik (tersedia dan dapat diakses oleh semua pelanggan layanan), dan *cloud hybrid* (kombinasi *cloud* publik dan privat). *Cloud* pribadi secara inheren dapat dipercaya dan aman tetapi ada beberapa batasan yang menghambat *cloud* pribadi untuk pemrosesan dan penyimpanan big data. Keterbatasan pertama adalah skalabilitas. Membangun cloud pribadi yang sangat terukur membutuhkan investasi modal yang besar. Hal ini menjadi sangat sulit untuk secara akurat merencanakan kapasitas *cloud* pribadi ketika volume, kecepatan, dan variasi data terus berubah. Keterbatasan kedua adalah tidak tersedianya model analitis dan kerangka kerja perangkat lunak yang diperlukan untuk mengelola data yang heterogen. Keterbatasan ketiga adalah berbagi data. Terkadang, berbagi data harus tersedia di antara kolaborator resmi yang tidak memiliki akses atau tinggal di luar *cloud* pribadi namun karena masalah keamanan seringkali tidak selalu memungkinkan. Di sisi lain, *cloud* publik mendukung skalabilitas dan berbagi data dengan mudah. *Cloud* publik lebih rentan terhadap serangan keamanan dan privasi karena multi-tenancy mesin virtual dan data. *Cloud* hybrid adalah kombinasi *cloud* publik dan *cloud* pribadi. Ini menyatukan fitur-fitur yang melekat dari

cloud publik yaitu, skalabilitas, kekuatan pemrosesan dll. C*loud* pribadi yaitu, keamanan dan memberikan peluang untuk pemrosesan dan penyimpanan *big data*. C*loud hybrid* telah digunakan untuk menjaga privasi pemrosesan dan penyimpanan big data.

Ketika *cloud computing* digunakan untuk penyimpanan *big data*, pemilik data kehilangan kendali atas data. *Data outsourcing* berisiko karena *server cloud* mungkin tidak sepenuhnya dipercaya. Pemilik data harus sangat yakin bahwa *cloud* menyimpan data dengan benar sesuai dengan kontrak layanan. Salah satu cara untuk memastikan privasi pengguna *cloud* adalah dengan menyediakan mekanisme bagi sistem untuk memungkinkan pemilik data memverifikasi bahwa datanya yang tersimpan di *cloud* masih utuh. Oleh karena itu verifikasi integritas data sangat penting. Berikut ini akan dibahas kerangka verifikasi integritas, diikuti oleh skema verifikasi integritas populer untuk data dinamis. Perhatikan bahwa data di sebagian besar aplikasi *big data* bersifat dinamis.

Pemilik data dapat melakukan verifikasi integritas sendiri atau mendelegasikan tugas kepada pihak ketiga yang tepercaya. Kerangka dasar skema verifikasi integritas terdiri dari tiga pihak yang berpartisipasi: *klien, server* penyimpanan *cloud* (CSS) dan auditor pihak ketiga (TPA). Klien menyimpan data di cloud dan tujuan TPA adalah untuk memverifikasi integritas data. Siklus hidup utama dari skema verifikasi integritas jarak jauh terdiri dari langkah-langkah berikut.

1. Pengaturan dan unggah data: Untuk memverifikasi data tanpa mengambil file yang sebenarnya, klien perlu menyiapkan metadata verifikasi. Metadata dihitung dari data asli dan disimpan di samping data asli. Untuk penggunaan praktis, metadata harus berukuran lebih kecil dibandingkan dengan dataset asli. Metadata dihitung dengan bantuan *homomorphic linear authenticator* (HLA) atau Homomorphic verifiable tag (HVT). HLA atau HVA telah berevolusi dari tanda tangan digital seperti RSA dan BLS (skema matematika untuk memverifikasi

integritas data). Setiap blok yang disimpan di cloud disertai dengan tag HVT atau HLA. Metode verifikasi integritas saat ini juga menggunakan struktur data yang diautentikasi seperti *Merkel Hash Tree* (MHT) [30]. MHT mirip dengan pohon biner, setiap *node* akan memiliki maksimum dua simpul anak. MHT adalah pohon *hash* di mana daun adalah *hash* blok data.

2. Otorisasi untuk TPA: TPA yang dapat memverifikasi data dari server cloud atas nama pemilik data perlu disahkan oleh pemilik data. Ada juga risiko keamanan jika pihak ketiga dapat meminta bukti integritas yang tidak terbatas pada dataset tertentu. Langkah ini hanya diperlukan ketika klien ingin beberapa pihak ketiga memverifikasi data.
3. Tantangan dan verifikasi penyimpanan data: Untuk memverifikasi integritas data, pesan tantangan dikirim ke server oleh TPA atas nama klien. Server akan menghitung respons berdasarkan pesan tantangan dan mengirimkannya ke TPA. TPA kemudian dapat memverifikasi respons untuk menemukan apakah data tersebut utuh. Skema memiliki verifikasi publik jika verifikasi ini dapat dilakukan tanpa kunci rahasia klien. Sebagian besar skema, seperti pemrosesan data yang dapat dibuktikan (PDP) dan *proof of retrievability* (POR), mendukung verifikasi data publik. Masalah utama dengan skema verifikasi publik adalah bahwa hal itu dapat mengaktifkan praktik jahat. Misalnya, pesan tantangan sangat sederhana dan semua orang dapat mengirim pesan tantangan ke CSS untuk bukti blok file tertentu. Pengguna yang jahat dapat meluncurkan serangan penolakan layanan terdistribusi (DDOS) dengan mengirim beberapa tantangan dari banyak klien dengan menyebabkan *overhead* dan kemacetan tambahan dalam lalu lintas jaringan.

4. Pembaruan data: Pembaruan data terjadi ketika beberapa operasi dilakukan pada data. Klien perlu melakukan pembaruan untuk beberapa penyimpanan data cloud. Pembaruan data yang umum dapat mencakup memasukkan, menghapus, dan memodifikasi operasi.
5. Pembaruan metadata: Setelah beberapa operasi pembaruan dilakukan pada data, klien perlu memperbarui metadata (HLA atau HVT) sesuai dengan kunci yang ada. Metadata diperbarui untuk menjaga penyimpanan data dapat diverifikasi tanpa mengambil semua data.
6. Verifikasi data yang diperbarui: Klien juga perlu memverifikasi apakah pembaruan data diproses dengan benar atau tidak karena *cloud* tidak dapat sepenuhnya dipercaya. Ini adalah langkah penting untuk memastikan bahwa data yang diperbarui masih dapat diverifikasi dengan benar di masa mendatang.

C. Perlindungan Privasi Bagian Pemrosesan Data

Perlindungan privasi di bagian pemrosesan data dapat dibagi menjadi dua fase. Pada fase pertama bertujuan untuk melindungi informasi dari pengungkapan yang tidak diminta karena data yang dikumpulkan dapat berisi informasi sensitif tentang pemilik data. Pada fase kedua bertujuan untuk mengekstraksi informasi yang bermakna dari data tanpa melanggar privasi.

Ada beberapa model yang diusulkan untuk menangani masalah di atas. Beberapa dari mereka termasuk anonimitas untuk mencegah hubungan rekaman, keberagaman untuk mencegah hubungan atribut dan hubungan catatan, kedekatan untuk mencegah serangan probabilistik dan hubungan atribut misalnya dengan cara teknik anonimisasi. Sebelum diterbitkan, tabel asli dimodifikasi sesuai dengan persyaratan privasi yang ditentukan. Untuk menjaga privasi, salah satu operasi anonimisasi berikut diterapkan pada data.

A. Generalisasi: Generalisasi berfungsi dengan mengganti nilai atribut QID tertentu dengan deskripsi yang kurang spesifik. Dalam operasi ini beberapa nilai digantikan oleh nilai induk dalam taksonomi atribut. Contohnya dapat mewakili atribut pekerjaan dengan artis, bukan penyanyi atau aktor. Jenis-jenis teknik generalisasi meliputi generalisasi domain, generalisasi *subtree*, generalisasi multidimensi, generalisasi sejenis.
B. Penindasan: Dalam penindasan, beberapa nilai diganti dengan karakter khusus (contoh ‘‘ * ’), yang menunjukkan bahwa nilai yang diganti tidak diungkapkan. Contoh skema penekanan termasuk penekanan catatan, penekanan nilai, dan penekanan sel.
C. Anatomisasi: Dalam metode ini, data pada QID dan SA dirilis dalam dua tabel terpisah. Satu tabel berisi pengidentifikasi kuasi dan tabel lainnya berisi atribut sensitif. Kedua tabel berisi satu atribut umum yang sering disebut GroupID. Grup yang sama akan memiliki nilai yang sama untuk GroupID yang ditautkan dengan nilai sensitif dalam grup.
D. Permutasi: Dalam permutasi, hubungan antara pengidentifikasi kuasi dan atribut yang peka terhadap angka adalah tidak terkait dengan mempartisi sekumpulan catatan ke dalam grup dan mencampur nilai-nilai sensitif mereka dalam setiap kelompok.
E. Perturbasi: Dalam perturbasi, nilai data asli diganti oleh beberapa nilai data sintetis, sehingga informasi statistik yang dihitung dari data yang dimodifikasi tidak berbeda secara signifikan dari informasi statistik yang dihitung dari data asli. Beberapa contoh termasuk menambahkan noise, bertukar data, dan menghasilkan data sintetis. Masalah dengan gangguan adalah bahwa catatan yang diterbitkan adalah sintetis dan tidak berarti apa-apa di dunia nyata dan karenanya tidak ada artinya bagi penerima. Mereka hanya

mempertahankan properti statistik yang dipilih secara eksplisit oleh penerbit.

Anonimisasi data tingkat tinggi menunjukkan bahwa privasi dilindungi dengan baik namun di sisi lain juga dapat mempengaruhi kegunaan data yang berarti bahwa nilai yang lebih sedikit dapat diekstraksi dari data. Oleh karena itu, menyeimbangkan pertukaran antara privasi dan utilitas sangat penting dalam aplikasi big data. Pengurangan dalam utilitas data diwakili oleh hilangnya informasi. Berbagai metode telah diusulkan dalam literatur untuk mengukur kehilangan informasi, beberapa contoh termasuk distorsi minimal (Huang, 2014), metrik kelaikan, metrik ukuran kelas ekivalensi rata-rata yang dinormalisasi dan metrik teori informasi (Wong, 2010). Untuk memecahkan masalah *trade-off* antara privasi dan utilitas, algoritma PPDP biasanya mengambil pendekatan serakah untuk mencapai *trade-off* yang tepat. Algoritma ini bekerja dengan menghasilkan beberapa tabel menggunakan metrik yang diberikan pelestarian privasi dan kehilangan informasi, yang semuanya memenuhi persyaratan model privasi tertentu selama proses anonimisasi.

Untuk mengekstrak informasi yang berguna dari *big data* tanpa melanggar privasi, teknik pelestarian data pelestarian privasi telah dikembangkan untuk mengidentifikasi pola dan tren dari data. Teknik-teknik tersebut tidak dapat langsung diterapkan ke big data karena big data dapat berisi data yang besar, kompleks, dan bervariasi secara dinamis. Untuk menangani big data dalam cara yang efisien, teknik tersebut harus dimodifikasi, atau beberapa teknik khusus harus digunakan. Selain itu, teknik-teknik yang dimodifikasi harus mengatasi masalah privasi.

Ada beberapa teknik yang diusulkan untuk menganalisis data dalam skala besar dan kompleks. Teknik-teknik ini dapat secara luas dikelompokkan ke dalam teknik-teknik berbasis aturan pengelompokan, klasifikasi dan asosiasi.

1. *Clustering* adalah salah satu teknik pemrosesan data populer karena kemampuannya menganalisis data yang tidak dikenal. Gagasan

mendasar di balik pengelompokan adalah untuk memisahkan data input tidak berlabel menjadi beberapa kelompok yang berbeda (Gionis, 2009). Algoritma pengelompokan konvensional membutuhkan data dalam format yang sama dan dimuat ke dalam satu unit pemrosesan tunggal, yang tidak cocok untuk pemrosesan *big data*. Banyak solusi (Xu, 2015), telah disajikan dalam dekade terakhir namun karena sifat dari *big data*, mereka memiliki beberapa kelemahan, di antaranya kompleksitas komputasi dan masalah privasi adalah masalah utama. Untuk menangani masalah kompleksitas komputasi, dalam (Xu et. al., 2009), Shirkhorshidi et al. memperkenalkan solusi pengambilan sampel dan pengurangan dimensi untuk pengelompokan mesin tunggal dan solusi paralel dan pengurangan peta untuk pengelompokan banyak mesin. Untuk meningkatkan efisiensi, dalam (Xu, 2015) diusulkan pemrosesan paralel berbasis *cloud computing*. Untuk membuat pengelompokan layak untuk set data yang sangat besar, Feldman et al. disajikan pendekatan pemrosesan paralel di mana set inti dibuat menggunakan konstruksi pohon. Dibandingkan dengan algoritma pengelompokan tradisional, waktu pemrosesan dan jumlah energi yang diperlukan berkurang secara signifikan. Berdasar pada semua metode ini, privasi adalah perhatian utama. Pelestarian privasi dalam pengelompokan adalah masalah yang menantang ketika data kompleks volume besar terlibat. Pada masa-masa awal, metode-metode berbasis transformasi data geometri *hybrid* diusulkan untuk melindungi privasi dalam pengelompokan (Elmiseryand, 2010). Metode ini mengubah atribut numerik dengan terjemahan, penskalaan, dan rotasi. Meskipun tingkat privasi tertentu dapat dicapai, utilitas data biasanya berkurang. Jadi metode ini secara praktis tidak layak. Oliveira dan Zaiane mengusulkan metode untuk data terpusat dengan

menggunakan reduksi dimensi dan representasi berbasis kesamaan objek. Hal ini dikarenakan metode ini dirancang khusus untuk data terpusat, metode ini tidak dapat digunakan dengan *big data* yang ada secara terpusat. Untuk meningkatkan efisiensi pengelompokan dalam data baru (tidak dikenal), pengelompokan pelestarian privasi berdasarkan model probabilitas terdistribusi diusulkan (Fahed, 2014).

2. Klasifikasi adalah teknik pengidentifikasian, di mana kelompok yang sudah ditentukan memiliki data input baru. Mirip dengan algoritma pengelompokan, algoritma klasifikasi secara tradisional dirancang untuk bekerja di lingkungan terpusat. Untuk mengatasi permintaan big data, algoritma klasifikasi tradisional dimodifikasi agar sesuai dengan lingkungan komputasi paralel. Misalnya, dalam (Tekon, 2013), algoritma klasifikasi dirancang untuk memproses data dengan dua cara. Algoritma ini, dikenal sebagai mengklasifikasikan atau mengirim untuk klasifikasi, baik mengklasifikasikan data sendiri atau meneruskan data input ke classifier lain. Ini efisien secara komputasi terutama ketika menangani data yang besar dan kompleks. Dalam algoritma klasifikasi novel lainnya, Rebentrost et al. (2013) mengusulkan mesin vektor dukungan berbasis kuantum untuk klasifikasi big data. Metode ini mengurangi kompleksitas komputasi dan data pelatihan yang diperlukan. Keterbatasan utama dari metode ini adalah teknologi perangkat keras yang belum matang dalam komputasi kuantum. Algoritma klasifikasi yang dikembangkan untuk big data dapat mencapai tingkat kinerja yang masuk akal, algoritma ini juga tidak terlalu memperhatikan privasi data. Algoritma penambangan data pelestarian privasi lain diusulkan menggunakan teknik rekonstruksi acak (Evfimievski, 2003). Operasi acak dalam algoritma melindungi privasi data asli melalui pengacakan data.

Simpulan

Penggunaan media sosial di Indonesia yang kian meningkat sehingga berkontribusi terhadap meningkatnya volume *big data*. Peningkatan *volume big* data harus diimbangi dengan adanya perlindungan privasi yang juga harus ditemukan solusinya karena pelanggaran privasi hingga saat ini masih terjadi. Beberapa hal yang dapat menjadi jalan keluar sebagai upaya untuk melindungi privasi *big data* dalam media sosial diantaranya adalah pertama perlindungan privasi pada saat data dibuat yaitu membatasi akses, penggunaan anti-virus, penggunaan *Soccetpuppet, MaskMe*. Kedua perlindungan privasi saat penyimpanan data dilakukan dengan cara melakukan enkripsi data dan verifikasi integritas yang dilakukan oleh klien, server penyimpanan dan auditor. Ketiga, perlindungan privasi saat pemrosesan data yang terdiri dari dua fase yaitu fase pertama bertujuan untuk melindungi data dari pengungkapan yang tidak diinginkan oleh misalnya Teknik anonimisasi dan fase kedua yang bertujuan untuk mengekstraksi informasi dari data tanpa melanggar privasi. Meskipun pada paparan diatas beberapa masalah mengenai privasi ditemukan beberapa, akan tetapi dunia teknologi kian berkembang sehingga ahli teknologi informasi dan komunikasi harus selalu meng-*upgrade* kemampuan dalam hal teknologi informasi untuk meminimalisir berbagai hal yang dapat mengurangi privasi pengguna media sosial.

Daftar Pustaka

Fahad *et al*., (2009). ‘‘A survey of clustering algorithms for big data: Taxonomy and empirical analysis,’’ *IEEE Trans. Emerg. Topics Comput.*, vol. 2, no. 3, pp. 267–279.

A.M. Elmiseryand, H.Fu (2010).‘‘Privacy preserving distributed learning clustering of healthcare
data using cryptography protocols,’’ in *Proc. IEEE 34th Annu. Comput. Softw. Appl. Conf. Workshops*, Jul. 2010, pp. 140–145.

Evfimievski, J. Gehrke, and R. Srikant (2003). ‘‘Limiting privacy breaches in privacy preserving data mining,’’ in *Proc. ACM Symp. Principles Database Syst.*, pp. 211–222.

Gionis and T. Tassa (2009) ''anonymization with minimal loss of information,'' *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 2, pp. 206–219.

Allen, A. L. 2013. An ethical duty to protect one's own information privacy? Alabama Law Review 64 (4):845–866.

Boyd, d. m., & Ellison, N. B. (2007). Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication*, 13, 210–230. doi:10.1111/j.1083-6101.2007.00393.x

C. Hongbing, R.Chunming, H.Kai, W.Weihong & L.Yanyan (2015). ''Secure big data storage and sharing scheme for cloud tenants,'' *China Community* vol. 12, no. 6, pp. 106–115.

C. Tekin and M. van der Schaar (2013). ''Distributed online Big Data classification using context information,''in *Proc. Int. Conf. Commun., Control, Comput.* pp. 1435–1442.

D. Feldman, M .Schmidt,and C.Sohler,''Turning big data into tiny data: Constant-size coresets for k-means, PCA and projective clustering,'' in *Proc. ACM-SIAM Symp. Discrete Algorithms*, 2013, pp. 1434–1453.

Franedya, Roy (2020). Duh! 235 Juta Data Pengguna Instagram, YouTube & TikTok Bocor. Retrived from https://www.cnbcindonesia.com/tech/20200826143301-37-182138/duh-235-juta-data-pengguna-instagram-youtube-tiktok-bocor

H. Hu, Y. Wen, T.-S. Chua, and X. Li (2014). ''Toward scalable systems for big data analytics: A technology tutorial,'' *IEEE Access*, vol. 2, pp. 652–687.

Howard, P. N., & Parks, M. R. (2012). Social media and political change: Capacity, constraint, and consequence. *Journal of Communication*, 62, 359–362 doi:10.1111/j.1460-2466.2012.01626.x

K. Liang,W.Susilo, & J.K.Liu. (2015). ''Privacy-preserving ciphertext multi-sharing control for big data storage,'' *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 8, pp. 1578–1589.

K. Yang, X. Jia, K. Ren, B. Zhang, and R. Xie (2013). ''DAC-MACS: Effective data access control for multi-authority cloud storage systems,'' *IEEE Trans. Inf. Forensics Security*, vol. 8, no. 11, pp. 1790–1800.

K. Yang, X. Jia, and K. Ren (2015). ''Secure and verifiable policy update outsourcing for big data access control in the cloud,'' *IEEE Trans. Parallel Distrib. Syst.*, vol. 26, no. 12, pp. 3461–3470.

Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of social media. *Business Horizons*, 53, 59–68. doi:10.1016/j.bushor.2009.09.003

Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of social media. *Business Horizons*, 53, 59–68. doi:10.1016/j.bushor.2009.09.003

Kent, M. L. (2010). Directions in social media for professionals and scholars. In R. L. Heath (Ed.), *Handbook of public relations* (2nd ed., pp. 643–656). Thousand Oaks, CA: Sage.

L. Xu, C.Jiang, J.Wang,J.Yuan, & Y.Ren (2014).''Information security in big data: Privacy and
data mining,'' in *IEEE Access*, vol. 2, pp. 1149–1176.

L. Xu, C. Jiang, Y. Chen, Y. Ren, and K. J. Ray Liu (2015). ''Privacy or utility in data collection? A contract theoretic approach,'' *IEEE J. Sel. Topics Signal Process.*, vol. 9, no. 7, pp. 1256–1269.

Laney, D. (2001). Application Delivery Strategies. Meta Group, Retrieved from http://blogs.gartner.com/douglaney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf

Lewis, B. K. (2010). Social media and strategic communication: Attitudes and perceptions among college students. *Public Relations Journal*, 4(3), 1–23.

M. Green and G. Ateniese. (2012). ''Identity-based proxy re-encryption,'' in *Proc. Int. Conf. Appl. Cryptogr. Netw. Secur.*, 2007, vol. 4521. pp. 288–306. J.Shao,''Anonymous ID-based proxy re-encryption,''in *Proc.Int.Conf. Inf. Secur. Privacy*, vol. 7372. pp. 364–375.

N.Cao, C.Wang, M.Li, K.Ren, & W.Lou (2014).''Privacy-preserving multi-keyword ranked search over encrypted cloud data,'' *IEEE Trans. Parallel Distrib. Syst.*,vol. 25, no. 1, pp. 222–233.

Nissenbaum, H. 2010. Privacy in context: Technology, policy, and the integrity of social life.
Stanford, CA: Stanford University Press.

O.M.Soundararajan, Y.Jenifer, S.Dhivya, & T.K.P. Rajagopal (2014).''Data security and privacy
in cloud using RC6 and SHA algorithms,'' *Netw. Commun. Eng.*, vol. 6, no. 5, pp. 202–204.

O'Reilly, T. (2005). What is Web 2.0: Design patterns and business models for the next generation of software.
O'Reilly Media. Retrieved from http://oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html

P. Mell and T. Grance. (2011). ''The NIST definition of cloud computing,'' *Nat. Inst. Standards Technology.*

Petronio, S. (2002). Boundaries of Privacy: Dialectics of Disclosure. USA: State University of New York

R. C. W. Wong and A. W.-C. Fu (2010). ''Privacy-preserving data publishing: An overview,'' *Synth. Lectures Data Manage.*, vol. 2, no. 1, pp. 1–138.

R. Xu and D. Wunsch. (2009). *Clustering*. New York, NY, USA: Wiley.

Russo, A., Watkins, J., Kelly, L., & Chan, S. (2008). Participatory communication with social media. *Curator: The Museum Journal*, 51, 21–31. doi:10.1111/j.2151-6952.2008.tb00292.x

Solove, D. J. 2008. Understanding privacy. Cambridge, MA: Harvard University Press.

Tavani, H. T. 2008. Informational privacy: Concepts, theories, and controversies. In The handbook of information and computer ethics, ed. K. E. Himma and H. T. Tavani, 131–164. Hoboken, NJ: Wiley.

Thom-Santelli, J., Millen, D. R., & DiMicco, J. M. (2010, August). *Characterizing global participation in an enterprise SNS*. Paper presented at the 3rd international conference on Intercultural collaboration, Copenhagen, Denmark.

U. Troppens, R. Erkens, W. Muller-Friedt, R. Wolafka, and N. Haustein (2014). *Storage Networks Explained: Basics and Application of Fibre Channel SAN, NAS, iSCSI, InfiniBand and FCoE*. New York, NY, USA: Wiley.

X. Huang and X. Du (2014). ''Achieving big data privacy via hybrid cloud,'' in *Proc. Int. Conf. INFOCOM*, pp. 512–517.

Z. Xiao and Y. Xiao (2013). ''Security and privacy in cloud computing,'' *IEEE Commun. Surveys Tuts.*, vol. 15, no. 2, pp. 843–859.