

SURVEI LITERATUR INFORMATION SECURITY INTELLIGENT UNTUK MEMPREDIKSI POTENSI KERUSUHAN MELALUI ANALISA JARINGAN MEDIA SOSIAL

Munirul Ula

Sistem Informasi Universitas Malikussaleh Lhokseumawe

Jl. Cot Tgk Nie-Reulet, Aceh Utara, 141 Indonesia

email: nanggroe@gmail.com

Abstrak

Kajian ini bertujuan untuk menganalisa literature terdahulu yang berkaitan dengan penerapan Information Security Intelligent (ISI) untuk memprediksi potensi kerusakan melalui media sosial. Kajian literatur penelitian terdahulu menunjukkan bahwa berbagai platform media sosial di Internet seperti Twitter, Tumblr, Facebook, YouTube, Blog dan forum diskusi disalahgunakan oleh kelompok-kelompok ekstremis untuk menyebarkan kepercayaan dan ideologi mereka. Situs web microblogging populer seperti Twitter digunakan sebagai platform real time untuk berbagi informasi dan komunikasi selama perencanaan dan mobilisasi massa. Penerapan analisa jaringan media sosial untuk memprediksi kejadian kerusakan adalah area yang telah menarik perhatian beberapa peneliti selama beberapa tahun terakhir. Ada berbagai macam metode yang telah digunakan dalam literatur terkait prediksi kejadian kerusakan. Dalam jurnal ini, penulis melakukan kajian literatur mengenai semua metode yang ada dan melakukan analisis yang komprehensif untuk memahami situasi, tren dan kesenjangan penelitian. Analisa kajian ini menghasilkan karakterisasi, klasifikasi, dan meta-analisis dari puluhan jurnal untuk mendapatkan pemahaman yang lebih baik tentang literatur tentang potensi kejadian kerusakan dengan menggunakan metode sosial media intelligent.

Kata Kunci :Information Security Intelligent (ISI); Machine learning; Kejadian Kerusakan; Analisis Media Sosial

I.Pendahuluan

Beberapa tahun terakhir, media sosial telah muncul ke dalam bentuk dinamis komunikasi antar pribadi di seluruh dunia. Ini memfasilitasi pengguna untuk berbagi informasi secara konstan dan terus menerus, membuat koneksi dan menyampaikan pemikiran mereka di seluruh dunia melalui media yang berbeda. Misalnya- jejaring sosial (Facebook), micro-blogging (Twitter, Tumblr), berbagi gambar (Imgur, Flickr) dan hosting

video dan berbagi (YouTube, Dailymotion, Vimeo). Karena jangkauan yang tinggi dan popularitas situs web media sosial di seluruh dunia, organisasi menggunakan situs web ini untuk merencanakan dan memobilisasi acara untuk protes dan demonstrasi publik [1]. Studi tentang kerusuhan menunjukkan bahwa sebagian besar protes direncanakan dan dimobilisasi jauh sebelumnya [2] [3]. Kerumunan orang tentang protes ini di media sosial adalah sumber data untuk memprediksi kejadian kerusuhan. Secara tradisional, surat kabar telah digunakan sebagai sumber utama untuk analisis dan prediksi tersebut. Namun, kecepatan dan fleksibilitas publikasi pada platform media sosial mendapat perhatian dari berbagai organisasi untuk perencanaan dan membuat pengumuman dari berbagai protes, pemogokan, demonstrasi publik dan kerusuhan [1]. Para peneliti dari berbagai disiplin ilmu seperti psikologi, ilmu sosial dan ilmu komputer terus-menerus mengembangkan metode untuk membuat peringatan dini untuk peristiwa terkait kejadian kerusuhan.

Kehadiran konten perencanaan kejadian kerusuhan adalah keprihatinan utama bagi pemerintah dan lembaga penegak hukum [4]. Pemerintah dan lembaga penegak hukum harus mencari pengguna dan komunitas tersebut untuk menghentikan promosi kebencian dan menjaga kedamaian di negara tersebut. Oleh karena itu, deteksi otomatis dan analisis konten perencanaan protes di media sosial adalah masalah penelitian penting dalam domain ISI. Memantau keberadaan konten seperti itu di media sosial dan melacak informasi ini dalam waktu nyata adalah penting bagi analis keamanan yang bekerja untuk lembaga penegak hukum. Selama 10 tahun terakhir, sejak 2005 beberapa pendekatan, teknik, algoritma dan alat telah diusulkan untuk menambang data dan membawa solusi untuk masalah yang dihadapi oleh Information security Intelligent (ISI) [2] [5] [6].

Tujuan dari penelitian yang disajikan dalam artikel ilmiah ini adalah untuk melakukan survei literatur secara sistematis terhadap metode-metode yang telah digunakan sebelumnya dalam artikel ilmiah terdahulu. Tujuan penulis adalah melakukan analisis komprehensif atas artikel-artikel ini untuk memahami dengan lebih baik, kesenjangan penelitian, teknik, dan arah.

2.Kerangka Kajian Literatur

Kajian literature dimulai dengan membuat daftar Google Scholar yang mewakili topik atau bidang masalah terkait. Misalnya, beberapa Google Scholar pencarian untuk mengambil jurnal yang relevan adalah: "''; peringatan dini "''; prediksi kerusakan online '. Penulis mencari artikel yang relevan menggunakan Google Scholar [2] yang merupakan mesin pencari berbasis web yang terkenal dan banyak digunakan untuk menemukan literatur ilmiah. Penulis menggunakan Google Scholar karena memiliki cakupan dan indeks artikel yang baik dan juga menyediakan mekanisme yang kuat untuk mengeksplorasi karya terkait, ide dan penulis. Google Scholar juga menyediakan data tentang seberapa sering dan seberapa baru jurnal telah dikutip yang juga merupakan metrik yang berguna (menilai dampak) untuk melakukan survei literatur penulis. Data diambil dari Judul, Kata Kunci dan Abstrak dari artikel sebagai hasil dari pencarian berbasis Google Scholar serta melalui artikel terkait dan dikutip oleh link yang disediakan untuk setiap artikel dalam hasil pencarian. Penulis melakukan meta analysis pada artikel dan menentukan relevansi artikel dengan topik atau area fokus penulis. Penulis melakukan klasifikasi satu kelas (Rule Based Classifier) pada setiap artikel dan memeriksa apakah itu memenuhi ruang lingkup dan fokus yang ditentukan.

3. Peramalan Kejadian kerusakan

Hasil survei literatur menunjukkan bahwa ada total 8 publikasi di konferensi dan jurnal untuk bidang prediksi kejadian kerusakan dengan menambang konten media sosial. Survei juga menunjukkan bahwa jumlah artikel ilmiah yang diterbitkan pada tahun 2014 termasuk satu jurnal [7], 3 artikel ilmiah lengkap [2] [8] [9] dan satu artikel ilmiah pendek [3]. Selama periode 2010 hingga 2013, penulis hanya menemukan 2 publikasi satu setiap tahun sementara tidak ada publikasi pada 2011 dan 2012 [11] [12]. Dapat disimpulkan bahwa masalah peramalan kejadian kerusakan hanya baru-baru ini menarik perhatian para peneliti.

4. Karakterisasi dan Klasifikasi Artikel Berdasarkan Analisis Meta

Dalam bagian tinjauan pustaka ini, penulis menyajikan studi berbasis karakterisasi pada penelitian sebelumnya yang dilakukan di bidang prakiraan peristiwa terkait kejadian kerusakan. Penulis menganalisis setiap artikel ilmiah dan membuat daftar semua dimensi untuk menunjukkan statistik. Penulis juga menyajikan statistik literatur yang ada yang menggunakan platform media sosial sebagai sumber data untuk melakukan eksperimen. Setiap dimensi selanjutnya diklasifikasikan

dalam sub-kategori dan memiliki sifat mereka yang terkait dengannya. Dimensi ini adalah sebagai berikut:

- 1 Sumber Data
- 2 Teknik
- 3 Fitur
- 4 Evaluasi
- 5 Jenis Analisis

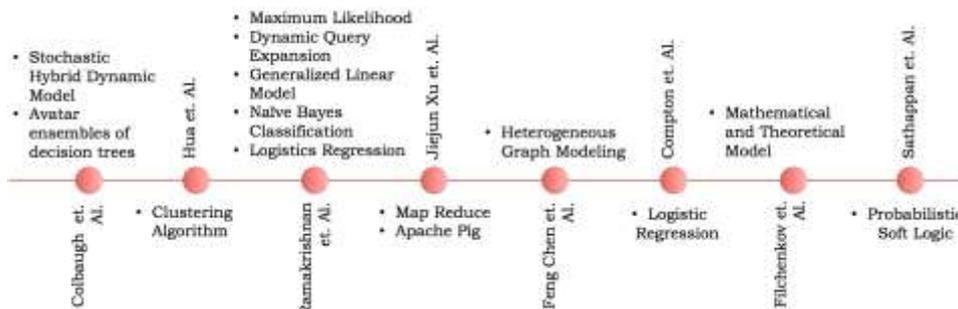
4.1 Situs Web Media Sosial Sebagai Sumber Data

Sebagaimana dibahas sebelumnya hanya ada 8 publikasi yang ada dalam domain peramalan kejadian kerusuhan. Oleh karena itu, untuk masalah peramalan acara untuk peristiwa terkait kejadian kerusuhan, penulis mengambil ID publikasi pada sumbu x dan sumber data pada sumbu y; warna mewakili platform media sosial yang berbeda dan sumber data lainnya. Hasil survei menunjukkan bahwa laporan media Twitter dan Berita adalah sumber data yang paling banyak digunakan untuk peramalan acara [1] [2] [9]. Terlepas dari artikel media berita, situs web micro-blogging adalah sumber informasi yang kaya. Penulis mengamati bahwa artikel ilmiah yang menggunakan sumber tunggal data menggunakan situs web mikro-blogging (Twitter dan Tumblr) [3] [7] [9]. Sebagian besar peneliti telah menggunakan berbagai sumber data (blog, forum, media berita, dan situs jejaring sosial lainnya) untuk pertukaran informasi [1] [2]. Penulis juga mengamati bahwa meskipun hosting video paling populer YouTube yang banyak digunakan belum digunakan dalam penelitian yang ada untuk perencanaan atau prediksi protes dan kerusuhan.

4.2 Teknik Data Mining dan Machine Learning

Gambar 1 mengilustrasikan pembelajaran mesin yang digunakan dalam prediksi kejadian kerusuhan. Hasil survei menunjukkan metode dan teknik yang digunakan dalam masing-masing artikel. Ramakrishnan [2] mengusulkan lima teknik berbeda untuk peramalan acara protes dan kerusuhan untuk lima jenis data dan model yang berbeda (berdasarkan volume, berdasarkan pendapat, pelacakan kegiatan, distribusi acara dan penyebab protes). Tidak seperti Ramakrishnan [2], penulis menemukan bahwa sebagian besar peneliti telah menerapkan pembelajaran ensemble pada beberapa data mining dan teknik

pembelajaran mesin untuk mencapai akurasi yang lebih baik dalam prediksi [1] [8] [11].



Gambar 1: Machine Learning Yang Digunakan Untuk Prediksi Kerusakan

Clustering, Regresi Logistik dan Ekspansi Kueri Dinamis adalah teknik yang umum digunakan untuk mengantisipasi peristiwa yang akan datang terkait dengan kejadian kerusakan atau protes. Penulis pergi melalui Bagian metodologi setiap artikel ilmiah dan mengamati bahwa pengakuan entitas bernama adalah fase umum untuk semua pendekatan yang diilustrasikan dalam artikel ilmiah yang dianalisis. Dalam pengakuan entitas bernama, mereka mengekstraksi beberapa entitas yang hadir dalam metadata kontekstual (tweet, komentar Facebook, Berita artikel) [13]. Entitas-entitas ini bisa berupa ekspresi spasial temporal, topik yang sedang dibahas posting. Daftar semua entitas dan kata kunci diperluas secara dinamis dengan menggunakan metode Ekspansi Kueri Dinamis di mana mereka menemukan kata kunci / entitas yang serupa dan relevan menggunakan sumber leksikal eksternal (misalnya- WordNet [4], VerbOcean [14]). Dynamic Query Expansion adalah proses berulang dan konvergen setelah kata kunci stabil. Mereka selanjutnya melakukan beberapa teknik pengelompokan dan klasifikasi pada entitas dan teks ini untuk memprediksi acara mendatang.

Tabel 1: Pengkategorian Literatur Tentang Peramalan Kerusakan

Dimensi	Kategori	Deskripsi
Fiture	Temporal	Kehadiran ekspresi terkait waktu
	Spatial	Kehadiran ekspresi berbasis lokasi
	Topic	Kehadiran topik terkait atau ekspresi terkait domain
	Content	Tambang teks untuk mengekstrak informasi terkait acara
	Demographic	Metadata berdasarkan demografis dan

		statistik lainnya
Evaluation	K-cross validation	mengoptimalkan output peramalan dengan memecah data menjadi k-sampel
	Precision	Mengevaluasi ketepatan hasil perkiraan
	Recall	Mengevaluasi kelengkapan hasil
	NA	Tidak ada teknik evaluasi yang disebutkan
Analisis	Content	Menambang hanya konten tekstual untuk ekstraksi fitur dan peramalan acara
	Community	Menambang profil pengguna dan jaringan untuk mengekstraksi informasi terkait acara

Teknik populer lain yang digunakan untuk peramalan acara protes dan kerusuhan adalah pemodelan grafik. Feng Chen et.al. [9] menggunakan pemodelan grafik heterogen sebagai pengayaan dan pra-pemrosesan kata kunci. Untuk mencapai hasil yang akurat dalam peramalan, mereka hanya menggunakan entitas yang difilter untuk pendekatan NPHGS. Menurut Feng Chen et.al. [9], grafik heterogen didefinisikan sebagai jaringan yang terdiri dari node, tepi dan hubungan di mana node adalah entitas yang diekstraksi menggunakan pengakuan entitas bernama. Mungkin ada beberapa jenis node yang setara dengan jumlah entitas yang diekstraksi (topik, temporal, spasial, organisasi dll). Tepi adalah tautan antara dua entitas dan relasi menentukan vektor fitur antara dua entitas. Misalnya, satu hubungan antara pengguna Twitter U (entitas: orang) dan istilah T (entitas: topik) dapat berupa jumlah tweet oleh pengguna U pada topik T. Perusahaan yang memiliki relevansi tinggi dan skor polaritas di atas ambang tertentu disaring dan digunakan dalam fase NPHSG berikutnya.

4.3 Karakterisasi dan Klasifikasi Artikel Prediksi Kejadian kerusuhan

Tabel 1 mengilustrasikan berbagai aspek dan sifat terkait yang digunakan oleh penulis untuk melakukan karakterisasi mendalam dari literatur yang ada. Fitur yang disebutkan dalam Tabel 1 terutama dari tiga jenis: entitas, analisis konten dan informasi demografis. Entitas diekstraksi menggunakan pengakuan entitas yang dinamai dan telah ditentukan

sebelumnya untuk masalah spesifik domain. Entitas-entitas ini adalah ekspresi temporal (hari ini, besok, 10 malam), ekspresi lokasi spasial (AS, di depan gedung putih, geocode) dan topik yang dibahas dalam posting (migrasi, protes). Analisis konten mencakup penambahan dan analisis metadata kontekstual. Misalnya, keberadaan tagar, @user disebutkan dalam tweet. Informasi demografis adalah ukuran berbasis statistik. Misalnya- jumlah tweet yang diposting oleh pengguna, usia dan lokasi pengguna. 75% artikel ilmiah telah menggunakan fitur spatiotemporal sebagai fitur diskriminatif untuk memprediksi kejadian. Sedangkan sisanya dari 25% hanya menggunakan timeline sebagai fitur diskriminatif dan menggunakan peramalan kejadian. Semua artikel ilmiah yang ada menggunakan fitur spatiotemporal juga menggunakan analisis konten untuk mengekstrak informasi yang lebih relevan dari teks mentah. Fitur analisis konten mencakup pencarian istilah kunci khusus domain.

Precision, Recall, dan validasi K-cross adalah teknik pengambilan informasi standar untuk mengevaluasi kinerja teknik yang diusulkan. Tabel 1 menunjukkan ukuran evaluasi umum yang digunakan oleh para peneliti. Precision menghitung 'ketepatan' dari metode yang diusulkan sementara penarikan kembali menghitung 'kelengkapan' dari pendekatan solusi. Validasi K-cross membagi testbed menjadi bagian K dan mengevaluasi model secara iteratif untuk setiap bagian. Validasi K-cross digunakan untuk mengoptimalkan hasil. 65% dari penelitian menggunakan ukuran Presisi untuk mengevaluasi kinerja pendekatan yang diusulkan di antaranya 25% artikel ilmiah menggunakan ukuran recall dan presisi. Sementara dalam 2 artikel ilmiah tidak ada metode evaluasi yang didefinisikan dan hanya dalam 1 artikel ilmiah, penulis menggunakan metode validasi k-cross untuk mengukur efektivitas pendekatan mereka

5. KESIMPULAN

Menerapkan information security intelligent dan analisis media sosial untuk memprediksi ancaman kejadian protes dan kerusuhan adalah area yang telah menarik perhatian beberapa peneliti selama 10 tahun terakhir. Penulis mengamati peningkatan dalam penelitian di Indonesia selama beberapa tahun terakhir pada topik solusi untuk mengidentifikasi dan memperkirakan kejadian kerusuhan dan mobilisasi dengan menambang konten tekstual di media sosial. Analisis penulis menunjukkan bahwa situs web micro-blogging seperti Twitter dan Tumblr adalah dua sumber paling umum dari data media sosial untuk deteksi kejadian kerusuhan dan aplikasi peramalan. Penulis percaya bahwa Twitter telah

sangat berperan dalam memfasilitasi pergerakan politik dibandingkan dengan platform media sosial lainnya karena karakteristik inherennya dalam berbagi teks pendek melalui pesan langsung dan hubungan pengikut. Sangat menarik untuk diamati bahwa meskipun popularitas dan penetrasi YouTube yang luar biasa sebagai situs berbagi video online, ini belum digunakan dalam penelitian yang ada untuk perencanaan atau prediksi protes dan kerusuhan.

Analisis penulis menunjukkan ada berbagai sumber informasi dan metode berbasis pembelajaran mesin yang digunakan oleh para peneliti untuk prediksi kejadian kerusuhan Clustering, Regresi Logistik dan Ekspansi Kueri Dinamis adalah teknik yang umum digunakan untuk memprediksi peristiwa yang akan datang terkait dengan kerusuhan atau protes. Named Entity Recognition (NER) adalah komponen umum dalam pipa pemrosesan teks untuk berbagai pendekatan dan teknik yang diusulkan. Pemodelan grafik juga merupakan teknik yang diadopsi oleh beberapa peneliti untuk masalah peramalan acara protes. Analisis penulis menunjukkan metadata penambangan kontekstual dari sebuah posting dan informasi spasial temporal hadir dalam konten adalah fitur yang paling umum digunakan untuk memprediksi peristiwa terkait kejadian kerusuhan. Dalam studi yang ada, penulis menemukan bahwa peneliti maksimum mengevaluasi keakuratan pendekatan prediksi mereka dengan menghitung ketepatan hasil mereka.

DAFTAR PUSTAKA

1. Muthiah, S., Huang, B., Arredondo, J., Mares, D., Getoor, L., Katz, G., Ramakrishnan, N.: Planned Protest Modeling in News and Social Media
2. Ramakrishnan, N., Butler, P., Muthiah, S., Self, N., Khandpur, R., Saraf, P., Wang, W., Cadena, J., Vullikanti, A., Korkmaz, G., Kuhlman, C., Marathe, A., Zhao, L., Hua, T., Chen, F., Lu, C.T., Huang, B., Srinivasan, A., Trinh, K., Getoor, L., Katz, G., Doyle, A., Ackermann, C., Zavorin, I., Ford, J., Summers, K., Fayed, Y., Arredondo, J., Gupta, D., Mares, D.: 'beating the news' with embers: Forecasting civil unrest using open source indicators. In: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD '14, pp. 1799–1808. ACM, New York, NY, USA (2014).

3. Filchenkov, A.A., Azarov, A.A., Abramov, M.V.: What is more predictable in social media: Election outcome or protest action? In: Proceedings of the 2014 Conference on Electronic Governance and Open Society: Challenges in Eurasia. EGOSE '14, pp. 157–161. ACM, New York, NY, USA (2014).
4. Chris Hale, W.: Extremism on the world wide web: a research review. *Criminal Justice Studies* 25(4), 343–356 (2012)
5. Wang, M.C.G.A., Chen, X.Z.H., Mao, D.Z.W.: *Intelligence and security informatics* (2011)
6. Yannakogeorgos, P.: Rethinking the threat of cyberterrorism. In: Chen, T.M., Jarvis, L., Macdonald, S. (eds.) *Cyberterrorism*, pp. 43–62. Springer, ??? (2014).
7. Ula, M., & Fuadi, W. (2017, February). A method for evaluating information security governance (ISG) components in banking environment. In *Journal of Physics: Conference Series* (Vol. 812, No. 1, p. 012031). IOP Publishing.
8. Compton, R., Lee, C., Xu, J., Artieda-Moncada, L., Lu, T.-C., Silva, L., Macy, M.: Using publicly visible social media to build detailed forecasts of civil unrest. *Security Informatics* 3(1) (2014).
9. Xu, J., Lu, T.-C., Compton, R., Allen, D.: Civil unrest prediction: A tumblr-based exploration. In: Kennedy, W., Agarwal, N., Yang, S. (eds.) *Social Computing, Behavioral-Cultural Modeling and Prediction 2014*. Lecture Notes in Computer Science, vol. 8393, pp. 403–411. Springer
10. Chen, F., Neill, D.B.: Non-parametric scan statistics for event detection and forecasting in heterogeneous social media graphs. In: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD '14, pp. 1166–1175. ACM, New York, NY, USA (2014).
11. Hua, T., Lu, C.-T., Ramakrishnan, N., Chen, F., Arredondo, J., Mares, D., Summers, K.: Analyzing civil unrest through social media. *Computer* 46(12), 80–84 (2013).
12. Colbaugh, R., Glass, K.: Early warning analysis for social diffusion events. *Security Informatics* 1(1), 1–26 (2012)

13. Leets, L.: Responses to internet hate sites: Is speech too free in cyberspace? *Communication Law & Policy* 6(2), 287-317 (2001)
14. Mahmood, S.: Online social networks: The overt and covert communication channels for terrorists and beyond. In: *Homeland Security (HST), 2012 IEEE Conference on Technologies For*, pp. 574-579 (2012). IEEE
15. Fisher, A.: How jihadist networks maintain a persistent online presence. *Perspectives on Terrorism* 9(3) (2015)