

IMPLEMENTASI *WEB SCRAPING* DALAM PENGUMPULAN BERITA KRIMINAL PADA MASA PANDEMI COVID-19

Studi Kasus: Situs Berita detik.com

(Implementation of Web Scraping in Criminal News Collection during Covid-19 Pandemic)

Salim Satriajati¹, Satria Bagus Panuntun², Setia Pramana^{1,2}

Program Studi Komputasi Statistik, Politeknik Statistika STIS¹
Subdirektorat Pengembangan Model Statistik, Badan Pusat Statistik RI²
Jalan Otto Iskandardinata No.64C, Jatinegara, Jakarta Timur,
Daerah Khusus Ibukota Jakarta 13330
221709994@stis.ac.id, 15.8887@stis.ac.id, setia.pramana@stis.ac.id

ABSTRAK

Saat ini telah banyak situs berita yang menyediakan informasi terkait kejadian maupun fenomena. Di sisi lain, pandemi Covid-19 memunculkan krisis dan masalah multidimensi. Salah satunya adalah timbulnya kriminalitas di tengah masyarakat. Penelitian ini bertujuan untuk mengumpulkan berita kriminal yang terjadi pada masa pandemi Covid-19 dari situs berita. Adapun pengumpulan informasi dari situs berita menggunakan teknik *web scraping*. *Web scraping* adalah suatu teknik penggalan informasi dari situs web. Berita yang berhasil dikumpulkan, kemudian dapat dianalisis mengenai adanya kemungkinan tren kejadian kriminal beriringan dengan tren pandemi Covid-19 di Indonesia. Situs berita yang digunakan pada penelitian ini adalah detik.com. Berdasarkan situs Alexa Internet (alexa.com), detik.com menjadi salah satu situs berita yang paling sering diakses dan masuk ke dalam 10 besar situs web dengan *traffic* tertinggi di Indonesia. Sedangkan data Covid-19 di Indonesia bersumber dari situs KawalCOVID19.id. Hasil penelitian menunjukkan bahwa jumlah berita kriminal dan jumlah kasus terkonfirmasi Covid-19 memiliki tren harian yang sama, yakni makin meningkat. Berdasarkan penelitian ini, dapat disimpulkan bahwa *Web scraping* dapat diimplementasikan untuk mengumpulkan berita. Hasil dari *web scraping* selanjutnya dapat digunakan untuk mengetahui tren jumlah berita kriminal harian yang kemudian dibandingkan dengan tren harian jumlah kasus terkonfirmasi Covid-19 di Indonesia.

Kata kunci: *web scraping, kriminal, covid-19*

ABSTRACT

Currently, there are many news sites that provide information related to events and phenomena. On the other hand, the Covid-19 pandemic caused a crisis and multidimensional problem. One of them is the emergence of crime in the community. This study aims to collect news on crimes that occurred during the Covid-19 pandemic from news sites. The gathering of information from news sites used web scraping techniques. Web scraping is a technique of extracting information from websites. The news that is collected can then be analyzed about possible trends in criminal events along with the trend of the Covid-19 pandemic in Indonesia. The news site used in this study is detik.com. Based on the Alexa Internet site (alexa.com), detik.com is one of the most frequently accessed news sites and is included in the top 10 websites with the highest traffic in Indonesia. Meanwhile, data for Covid-19 in Indonesia collected from KawalCOVID19.id website. The results showed that the number of criminal news reports and the number of confirmed cases of Covid-19 have the same daily trend, which is increasing. Based on this research, it can be argued that Web scraping can be implemented to gather news. The results of web scraping can then be used to find out the trend of the number of daily criminal news which is then compared with the daily trend of the number of confirmed cases of Covid-19 in Indonesia.

Keywords: *web scraping, crime, covid-19*

PENDAHULUAN

Saat ini, masyarakat Indonesia cukup menaruh perhatian terhadap isu maupun kejadian yang terjadi di tengah mereka. Kenyataan tersebut ternyata beriringan dengan beragamnya situs berita, terutama di era kemajuan teknologi informasi dan komunikasi seperti saat ini (Adli & Firgia, 2018). Situs berita menyediakan informasi terkait kejadian maupun fenomena di tengah masyarakat dalam suatu situs web secara dalam jaringan (daring), sehingga situs berita selalu tersedia setiap

saat. Adanya situs berita yang beragam membuat masyarakat dapat memperoleh informasi dengan cepat dan mudah.

Di sisi lain, timbulnya pandemi Covid-19 sepanjang tahun 2020 ini tidak hanya menimbulkan permasalahan kesehatan masyarakat di Indonesia dan dunia. Pandemi Covid-19 juga telah memunculkan permasalahan di berbagai bidang. Adanya pandemi membuat aktivitas perekonomian global melambat dan turut memicu perlambatan perekonomian di Indonesia (Nasution et al., 2020). Kebijakan pemerintah yang melakukan intervensi dalam meminimalisasi penyebaran virus Covid-19, seperti kebijakan Pembatasan Sosial Berskala Besar (PSBB) juga berdampak pada lemahnya perekonomian di Indonesia (Hanoatubun, 2020).

Ketidakpastian perekonomian di Indonesia di masa pandemi Covid-19 dikhawatirkan memunculkan permasalahan lain. Salah satunya adalah timbulnya kriminalitas di tengah masyarakat. Permasalahan ekonomi seperti kemiskinan ternyata memiliki pengaruh yang positif dan signifikan terhadap kriminalitas di Indonesia (Rahmalia & Triani, 2019). Di samping itu, dilansir oleh situs berita *mediaindonesia.com*, Kabid Humas Polda Metro Jaya, Kombes Yusri Yunus menyebutkan bahwa selama pandemi Covid-19, tepatnya pada rentang waktu Maret hingga Mei 2020, kejadian kriminal di wilayah hukum Polda Metro Jaya didominasi oleh empat kasus, yakni kejahatan jalanan (*street crime*), narkoba, penyebaran hoaks, dan ujaran kebencian. Tren kejahatan dari yang sebelumnya menyasar rumah, selama Covid-19 berpindah menjadi ke tempat usaha dan kejahatan berupa pembegalan.

Hal tersebut yang menjadi latar belakang diadakannya penelitian ini, yakni untuk mengumpulkan berita kriminal dari situs berita, khususnya yang terjadi pada masa pandemi Covid-19. Pengumpulan berita adalah salah satu penerapan dari *Big data*, dimana *Big Data* dapat menjadi salah satu sumber data alternatif untuk menghasilkan statistik resmi (*official statistics*) (Pramana et.al., 2017). Statistik resmi yang terkait dengan kriminal sendiri rutin dipublikasikan oleh Badan Pusat Statistik Republik Indonesia setiap tahunnya. Publikasi tersebut bernama "Statistik Kriminal". Selanjutnya, kemunculan berita kriminal pada situs berita juga dapat digunakan sebagai proksi (pendekatan) kejadian kriminal yang sesungguhnya terjadi dalam lingkungan masyarakat.

Pengumpulan berita di dalam situs berita pada penelitian ini menggunakan teknik *web scraping*. *Web scraping* adalah teknik untuk mendapatkan informasi dari situs web secara otomatis tanpa harus menyalinnya secara manual (A. Yani et al., 2019). Tujuan dari *web scraping* adalah untuk mencari informasi pada bagian tertentu. Tidak seperti kegiatan *web crawling* yang mengunjungi seluruh situs yang berhubungan dengan situs utamanya, kegiatan *web scraping* hanya melakukan ekstraksi data tertentu saja dari situs yang dituju sesuai dengan kebutuhan (Setiawan et al., 2020). Hasil dari *web scraping* sendiri dapat dimanfaatkan kembali oleh sistem lain maupun dianalisis lebih lanjut (Rahmatulloh & Gunawan, 2020).

Penelitian yang menerapkan teknik *web scraping* pada situs berita pernah dilakukan oleh Sonya & Prihandoko pada 2016 dengan melakukan *web scraping* pada tiga situs berita yakni detik.com, liputan6.com dan vivanews.com untuk mendapatkan data mengenai bencana alam.

Selain itu, penelitian mengenai penggunaan artikel berita sebagai proksi pernah dilakukan oleh George Willis dan Emmanouil Tranos pada tahun 2020 untuk menangkap volume minat di ranah publik tentang perusahaan penyedia transportasi di Amerika Serikat, yakni Uber dengan menggunakan artikel yang diterbitkan oleh situs berita NYTimes sebagai proksinya. Penggunaan proksi sendiri disebabkan karena keterbatasan akses data yang diberikan pihak Uber.

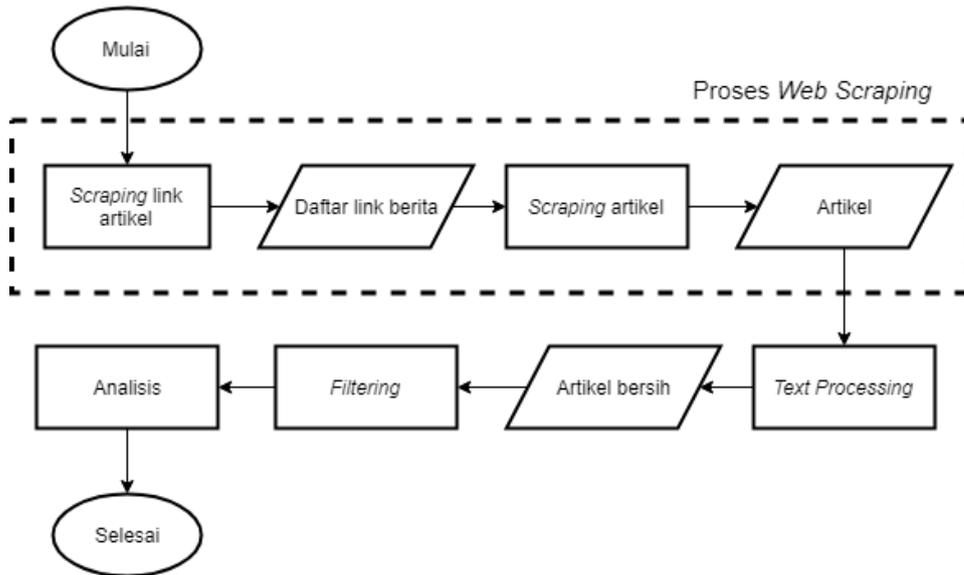
METODE

Situs berita yang digunakan pada penelitian ini adalah detik.com. Berdasarkan situs Alexa Internet (*alexa.com*), detik.com menjadi salah satu situs berita yang paling sering diakses dan masuk ke dalam 10 besar situs web dengan *traffic* tertinggi di Indonesia.

Pengumpulan berita kriminal dari situs detik.com dimulai pada tanggal 1 Januari 2020 dan dalam penelitian ini dibatasi sampai 31 Agustus 2020. Sedangkan data harian jumlah kasus terkonfirmasi Covid-19 di Indonesia yang digunakan sebagai pembanding dengan jumlah kasus kriminal bersumber dari situs KawalCOVID19.id dalam format *spreadsheet*.

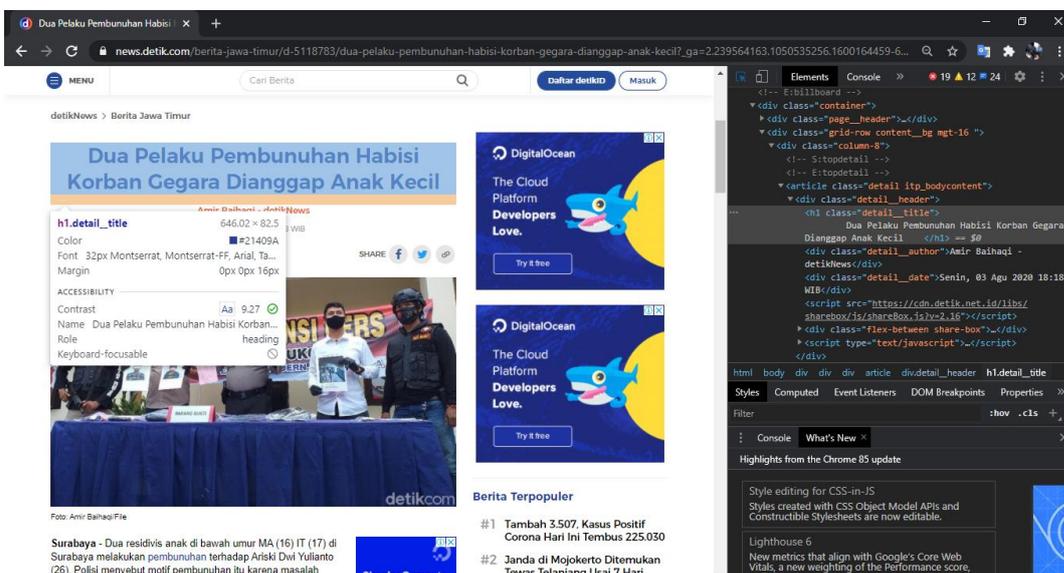
Dalam penelitian ini, alat yang digunakan untuk melakukan web scraping adalah bahasa pemrograman Python beserta *package scrapy*. *Package scrapy* dipilih karena scrapy bersifat *open source*, fleksibel dan mampu mengekstrak dokumen HTML dengan memilih bagian tertentu (Rohman et al., 2019). Program *web scraping* dijalankan pada aplikasi Visual Studio Code. Sedangkan untuk visualisasi data menggunakan aplikasi Tableau.

Data yang akan diekstrak terdiri dari isi berita dan beberapa informasi penyerta yakni judul berita, tanggal, tempat kejadian kriminal (tempat peliputan) dan tautan (*link*) berita. Adapun parameter dalam pengambilan informasi adalah berdasarkan hasil pencarian pada situs berita dengan menggunakan *keyword* (kata kunci) yang diinput (Sonya, 2016). Kata kunci yang digunakan dalam penelitian ini berupa jenis-jenis kriminal, yakni pembunuhan, penganiayaan, pencurian, narkoba, perampokan, pencabulan, penjangbretan, perampasan.



Gambar 1. Flowchart proses *web scraping* dan analisis

Adapun proses dari *web scraping* secara umum adalah sebagai berikut. Pertama, identifikasi terlebih dahulu tag HTML pada situs berita yang memuat berita dan informasi penyerta. Identifikasi dokumen HTML bertujuan untuk mengapit informasi yang akan diambil (Josi et al., 2014). Setelah identifikasi dilakukan, kemudian *web scraper* dapat dibuat. Web scraper dalam *package scrapy* dibentuk dalam file python berekstensi .py yang biasa disebut dengan istilah "spider".



bertempat di luar negeri dihapus. *Filtering* tempat kejadian secara manual juga dimaksudkan untuk memperbaiki kesalahan ketik pada nama tempat. Hasil *filtering* tempat kejadian tersebut menghasilkan artikel sebanyak 8.853 buah artikel.

Data hasil *filtering* tersebut, selanjutnya masuk ke dalam tahap analisis. Analisis yang digunakan adalah analisis deskriptif dengan melakukan pendeskripsian dan penggambaran terhadap data berita kriminalitas yang berhasil didapat. Dalam analisis, juga diteliti mengenai kemungkinan adanya tren jumlah kejadian kriminal beriringan dengan tren jumlah kasus terkonfirmasi Covid-19 di Indonesia.

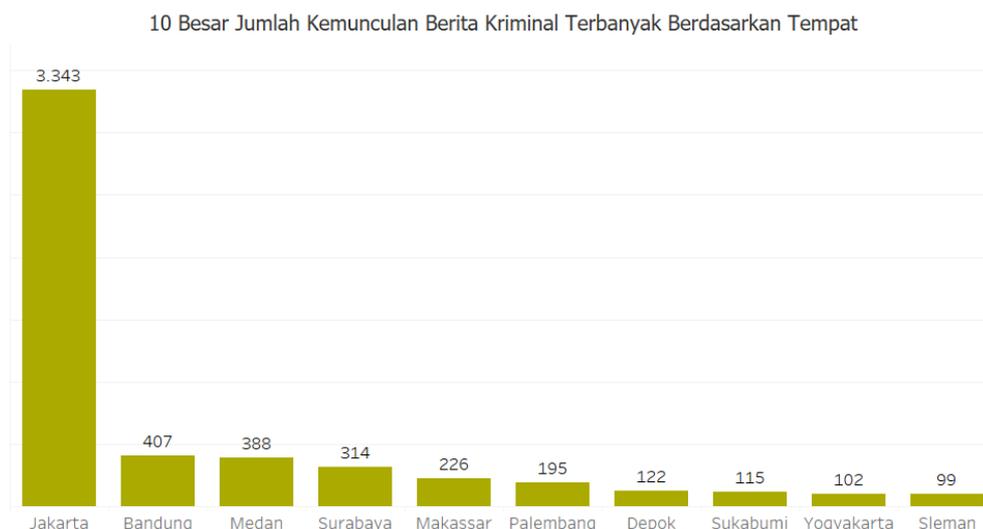
HASIL DAN PEMBAHASAN

Dari 8.853 buah artikel hasil dari proses *web scraping* hingga *filtering*, selanjutnya untuk mendapatkan *insight*, data artikel tersebut kemudian dianalisis lebih lanjut. Hal pertama yang dapat ditelusuri adalah mengenai jumlah kemunculan berita kriminal berdasarkan jenis kriminal, seperti yang terlihat pada grafik di bawah ini.



Gambar 5. Jumlah Kemunculan Berita Kriminal Berdasarkan Jenisnya

Berdasarkan **Gambar 5** di atas, dapat dilihat bahwa pembunuhan adalah jenis kriminal yang paling banyak muncul dalam situs detik.com dibanding jenis kriminal lainnya dengan jumlah kemunculan sebanyak 2.710 buah artikel, kemudian disusul penganiayaan dengan 1.714 buah artikel, pencurian dengan 1.669 buah artikel hingga jenis kriminal yang paling sedikit muncul yakni perampasan dengan 122 buah artikel. Lalu, selain berdasarkan jenis kriminal, dapat diketahui juga mengenai jumlah kemunculan berita kriminal berdasarkan tempat kejadian, seperti yang terlihat pada grafik di bawah ini.



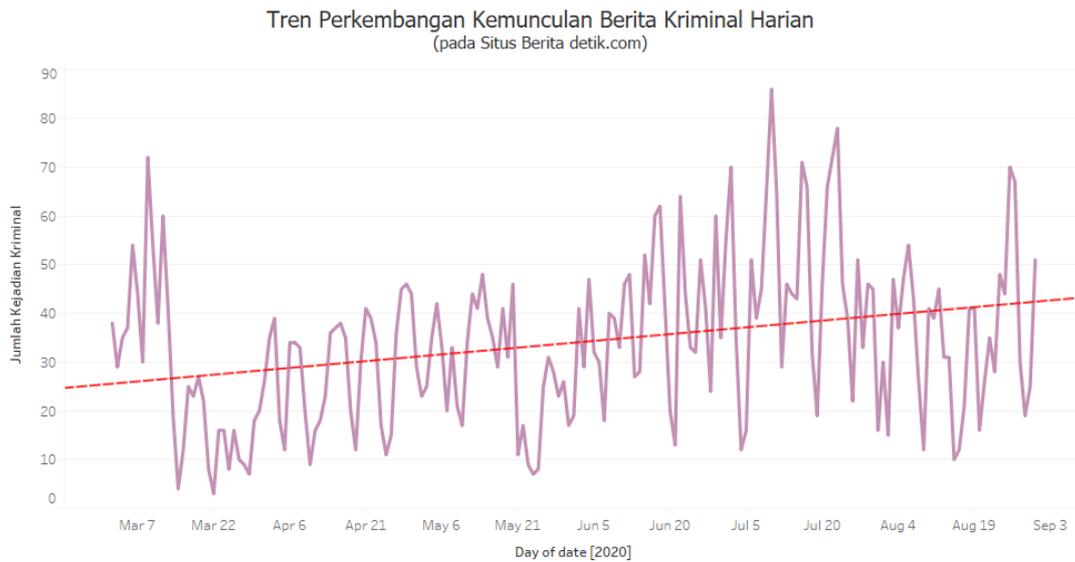
Gambar 6. 10 Besar Jumlah Kemunculan Berita Kriminal Terbanyak Berdasarkan Tempat

Gambar 6 merupakan grafik mengenai 10 tempat kejadian dengan jumlah kemunculan berita kriminal terbanyak pada situs detik.com, dimana Jakarta merupakan tempat yang paling banyak muncul yakni sebanyak 3.343 buah artikel. Kemudian disusul Bandung dengan 407 artikel dan Medan dengan 388 buah artikel. Sedangkan, artikel di luar 10 besar tersebut yakni sebanyak 3542 buah artikel, tersebar di beberapa lokasi di Indonesia. Data mengenai kemunculan berita kriminal berdasarkan tempat dan waktu kejadian selengkapnya dapat dilihat pada tabel di bawah ini.

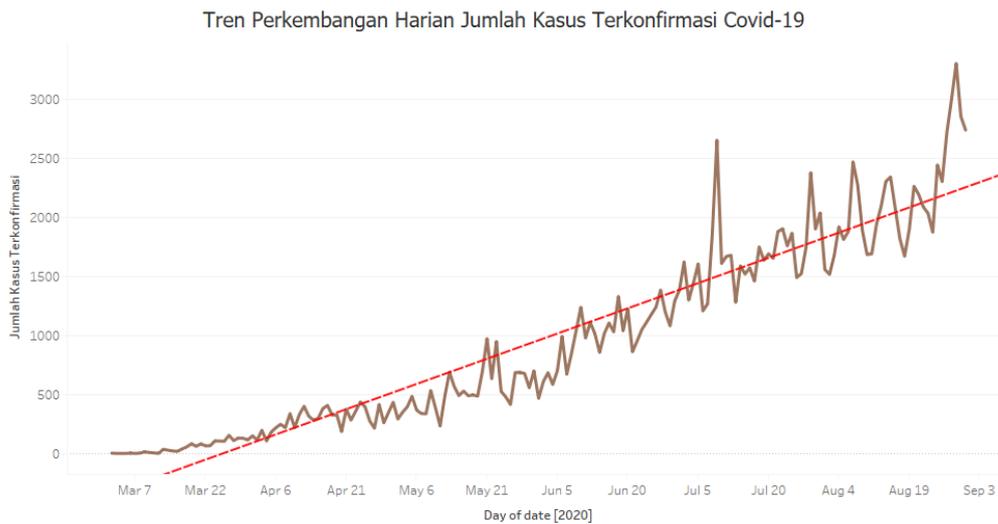
Tabel 1. Jumlah kemunculan berita kriminal berdasarkan tempat dan waktu kejadian

Daerah	Bulan								Jumlah
	Januari	Februari	Maret	April	Mei	Juni	Juli	Agustus	
Jakarta	433	410	260	302	334	563	618	423	3343
Bandung	47	82	51	57	35	30	72	33	407
Medan	99	47	42	30	50	33	64	23	388
Surabaya	56	62	68	22	14	29	34	29	314
Makassar	38	31	28	22	15	27	31	34	226
Palembang	19	17	34	28	10	28	49	10	195
Depok	6	0	3	38	9	36	13	17	122
Sukabumi	30	18	5	5	7	18	26	6	115
Yogyakarta	12	28	6	8	9	8	14	17	102
Sleman	25	11	19	7	7	7	8	15	99
Lainnya	565	607	291	334	378	372	520	475	3542
Jumlah	1330	1313	807	853	868	1151	1449	1082	8853

Insight selanjutnya yang dapat ditelusuri adalah mengenai keterkaitan antara pandemi Covid-19 dengan kemunculan berita kriminal pada masa tersebut. Keterkaitan antara keduanya dapat dilihat dengan menggambarkan tren kasus terkonfirmasi dan kemunculan berita secara harian. Kedua tren tersebut dapat dilihat pada **Gambar 7** dan **Gambar 8** di bawah ini.

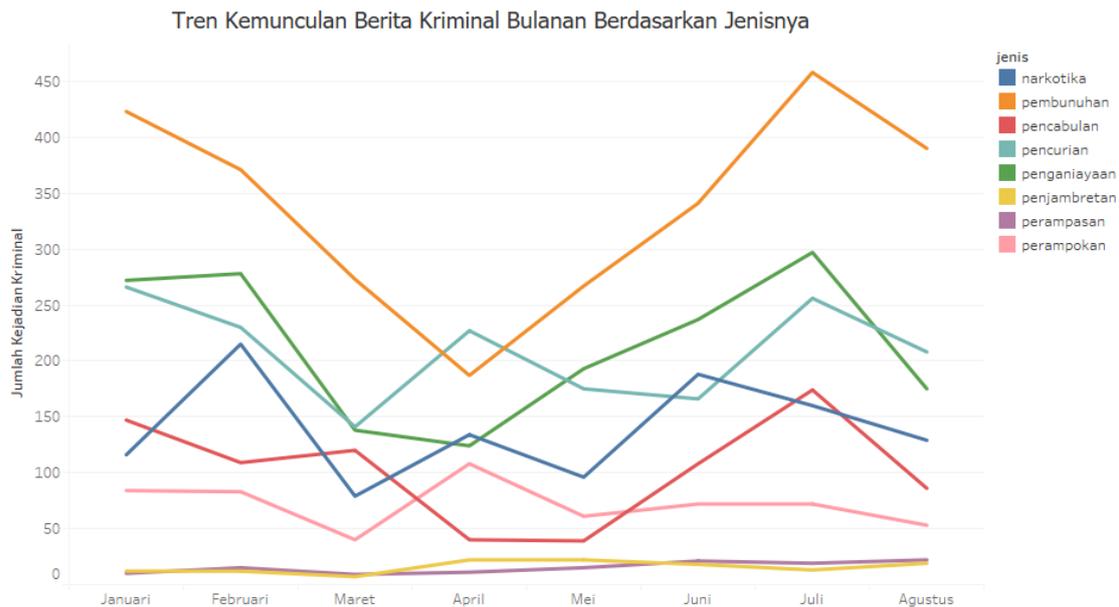


Gambar 7. Tren Perkembangan Kemunculan Berita Kriminal Harian



Gambar 8. Tren Perkembangan Harian Jumlah Kasus Terkonfirmasi Covid-19

Dapat dilihat bahwa pada **Gambar 7** mengenai tren kemunculan berita kriminal dan **Gambar 8** mengenai jumlah kasus terkonfirmasi Covid-19, keduanya memiliki tren harian yang sama, yakni makin meningkat. Namun, informasi yang dapat disimpulkan mengenai keterkaitan antara keduanya masih kurang jelas terlihat dan sulit diinterpretasikan. Maka dari itu, dilakukan penggambaran dengan mengakumulasikan kemunculan berita kriminal secara bulanan berdasarkan jenis kejahatan, seperti yang terlihat pada grafik di bawah ini.



Gambar 9. Tren Kemunculan Berita Kriminal Bulanan Berdasarkan Jenisnya

Berdasarkan **Gambar 9** di atas, dapat dilihat bahwa secara umum tren jumlah kemunculan berita kriminal pada seluruh jenis kriminal mengalami penurunan pada Januari sampai Februari, namun meningkat mulai pada bulan Maret dan terus meningkat sampai bulan Juli dan kemudian kembali menurun pada bulan Agustus.

Jika dikaitkan dengan adanya pandemi Covid-19 di Indonesia, dimana kasus terkonfirmasi Covid-19 di Indonesia pertama kali muncul pada 2 Maret 2020 dan kebijakan pemerintah untuk menanggulangi penyebaran virus Covid-19 yakni dengan pemberlakuan Pembatasan Sosial Berskala Besar (PSBB) mulai diberlakukan pada pertengahan April. Hal tersebut sejalan dengan tren jumlah kemunculan berita kriminal yang digambarkan pada **Gambar 9** di atas. Maka dapat diketahui bahwa kemunculan pandemi Covid-19 dan kebijakan PSBB berpotensi memiliki keterkaitan erat dengan kejadian kriminal yang terjadi di masyarakat. Namun, hal tersebut tentunya perlu diteliti lebih lanjut, terutama mengenai motif pelaku kejahatan, apakah pelaku kejahatan melakukan tindak kejahatan karena terdesak oleh keadaan maupun keterbatasan yang disebabkan pandemi Covid-19 atau tidak.

KESIMPULAN

Berdasarkan hasil di atas, maka dapat disimpulkan bahwa *Web scraping* dapat diimplementasikan pada situs berita, termasuk situs detik.com untuk mengumpulkan berita. Hasil dari *web scraping* selanjutnya dapat digunakan untuk mengetahui tren jumlah berita kriminal harian (sebagai proksi jumlah kasus kriminal sebenarnya) pada masa pandemi Covid-19 dan berpeluang melengkapi statistik resmi (*official statistic*) pada Statistik Kriminal melalui pendekatan *Big Data*.

Mengenai hasil analisis data berita kejahatan hasil *web scraping*, dapat disimpulkan bahwa tren kemunculan berita kejahatan pada situs detik.com secara umum meningkat setiap harinya, terutama setelah bulan Maret, dimana pada bulan tersebut terjadi kasus terkonfirmasi Covid-19 pertama kali di Indonesia dan bulan April dimana pada bulan tersebut pemerintah menjalankan kebijakan Pembatasan Sosial Berskala Besar (PSBB). Namun, keterkaitan antara munculnya tindak kejahatan dengan pandemi Covid-19 perlu diteliti lebih lanjut dengan melakukan penelusuran lapangan yang lebih mendalam, salah satunya adalah dengan melakukan pengidentifikasian mengenai motif pelaku kejahatan yang tidak dilakukan pada penelitian ini.

DAFTAR PUSTAKA

- Adli, M. A., & Firgia, L. (2018). Rancang Bangun Web Scraping Pada Media Online Berita Nasional. *Jurnal ENTER*, 1, 118–128.
- Ayani, D. D., Pratiwi, H. S., & Muhandi, H. (2019). Implementasi Web Scraping untuk Pengambilan Data pada Situs Marketplace. *Jurnal Sistem Dan Teknologi Informasi (JUSTIN)*, 7(4), 257. <https://doi.org/10.26418/justin.v7i4.30930>
- Hanoatubun, S. (2020). Dampak COVID-19 terhadap Perekonomian Indonesia. *EduPsyCouns Journal of Education, Psychology and Consulting*, 2(1), 146–153.
- Josi, A., Abdillah, L. A., & Suryayusra. (2014). Penerapan teknik web scraping pada mesin pencari artikel ilmiah. <http://arxiv.org/abs/1410.5777>
- Nasution, D. A. D., Erlina, E., & Muda, I. (2020). Dampak Pandemi COVID-19 terhadap Perekonomian Indonesia. *Jurnal Benefita*, 5(2), 212. <https://doi.org/10.22216/jbe.v5i2.5313>
- Nasution, D. A. D., Erlina, E., & Muda, I. (2020). Dampak Pandemi COVID-19 terhadap Perekonomian Indonesia. *Jurnal Benefita*, 5(2), 212. <https://doi.org/10.22216/jbe.v5i2.5313>
- Pramana, S., Yuniarto, B., Kurniawan, R., Yordani, R., Lee, J., Amin, I., Satyaning, P. P. N. L. P., Riyadi, Y., Hasyati, A. N., & Indriani, R. (2017). Big data for government policy: Potential implementations of bigdata for official statistics in Indonesia. *Proceedings - WBIS 2017: 2017 International Workshop on Big Data and Information Security*, 17–21. <https://doi.org/10.1109/IWBIS.2017.8275097>
- Rahmalia, S., & Triani, M. (2019). Pengaruh Tingkat Pendidikan, Pengangguran, dan Kemiskinan Terhadap Kriminalitas di Indonesia. *Jurnal Kajian Ekonomi Dan Pembangunan*, 1(1), 21–36. <https://doi.org/10.24036/jkep.v1i1.5345>
- Rahmatulloh, A., & Gunawan, R. (2020). Web Scraping with HTML DOM Method for Data Collection of Scientific Articles from Google Scholar. *Indonesian Journal of Information Systems*, 2(2), 16. <https://doi.org/10.24002/ijis.v2i2.3029>
- Rohman, M. S., Santoso, H. A., & Saraswati, G. W. (2019). Pemanfaatan Topic-Focused Crawler untuk Pembangunan Corpus Berita Bencana menggunakan Teknik Scrapy CSS Selector. *Seminar Nasional APTIKOM (SEMNASTIK) 2019*, 250–258.
- Setiawan, A., Santoso, L. W., & Adipranata, R. (2020). Klasifikasi Artikel Berita Bahasa Indonesia Dengan Naive Bayes Classifier. *Jurnal Infra*, 8(1), 146–151.
- Sonya, I. P., & Prihandoko. (2016). Analisis Web Scraping untuk Data Bencana Alam dengan Menggunakan Teknik Breadth-First Search Terhadap 3 Media Online. *Jurnal Ilmiah Informatika Komputer Universitas Gunadarma*, 21(3), 69–77.
- Willis, G., & Tranos, E. (2020). Using 'Big Data' to understand the impacts of Uber on taxis in New York City. *Travel Behaviour and Society*, 1–31.