

Iva Rachmawati
Hestutomo Restu Kuncoro
Dyah Lupita Sari
Sylvert Prian Tahalea



METODE KOMPUTASIONAL DALAM HUBUNGAN INTERNASIONAL

**METODE
KOMPUTASI**
DALAM HUBUNGAN INTERNASIONAL

deepublish / publisher

deepublish / publisher

METODE KOMPUTASI

DALAM HUBUNGAN INTERNASIONAL

**Iva Rachmawati
Hestutomo Restu Kuncoro
Dyah Lupita Sari
Sylvert Prian Tahalea**

LPPM UPNYK Press

LPPM UPNYK Press

Universitas Pembangunan Nasional 'Veteran' Yogyakarta
Rektorat Lantai 4, LPPM, Puslitbang
Jl. SWK 104 (Lingkar Utara), Condong Catur, Yogyakarta 55283
Telepon (0274) 486733, ext 154
Fax (0274) 486400
www.lppm.upnyk.ac.id
email: lppm@upnyk.ac.id

Cetakan Tahun 2023

Metode Komputasi dalam Hubungan Internasional
LPPM PRESS UPNYK

Jumlah halaman xiv + 112 hlm; 15.5x23 cm

ISBN:

Penata Letak : G.D. Ayu
Proofreader : Tiara Nabilah Azalia
Desain Sampul : Syaiful Anwar

Hak Cipta Dilindungi Undang-Undang Dilarang memperbanyak karya tulis ini dalam bentuk dan dengan cara apapun, termasuk fotokopi, tanpa izin tertulis dari penerbit.

Pengantar

Dalam dunia yang terus berubah dengan cepat, kompleksitas hubungan internasional semakin memunculkan dimensi-dimensi yang beragam. Dinamika politik, ekonomi, sosial, dan budaya antar negara tidak lagi dapat diartikan secara linear, melainkan sebagai rangkaian interaksi yang kompleks dan saling terkait. Guna menghadapi tantangan ini, diperlukan pendekatan analitis yang tidak hanya cerdas, tetapi juga mampu mengurai kerumitan tersebut dengan detail yang lebih mendalam. Inilah titik di mana metode komputasi memainkan peran kunci.

Metode komputasi, dengan penerapan teknik-teknik komputasi dan algoritma dalam analisis data, telah muncul sebagai alat paling bermanfaat untuk menjelajahi kompleksitas ini. Pada saat yang bersamaan, perkembangan teknologi telah menciptakan sebuah ekosistem di mana jumlah data yang tersedia semakin meluas dan mendalam. Data besar (*big data*) menjadi semakin mendominasi, mencakup informasi dari berbagai sumber, seperti media sosial, lalu lintas internet, data ekonomi, dan lebih banyak lagi. Fenomena ini menciptakan tantangan baru dalam pengolahan dan pemahaman informasi.

Metode analisis konvensional tidak lagi cukup untuk menghadapi volume data yang semakin membesar dan rumit ini. Inilah mengapa metode analisis yang memanfaatkan teknologi komputer muncul sebagai jawaban yang tepat. Metode ini memungkinkan kita untuk memanfaatkan kecepatan dan kapasitas komputasi untuk memproses data dalam skala yang tidak mungkin dicapai oleh analisis manual. Melalui algoritma-algoritma cerdas,

metode ini dapat mengidentifikasi pola-pola yang tersembunyi, hubungan yang kompleks, dan tren-tren yang sulit dilihat dalam tumpukan data besar.

Buku "Metode Komputasi dalam Hubungan Internasional" bertujuan untuk membuka pintu ke dunia baru ini bagi pembaca. Ini adalah panduan yang memberi pemahaman tentang bagaimana metode komputasi memungkinkan kita untuk memecah kerumitan hubungan internasional menjadi elemen-elemen yang dapat diuraikan. Dari analisis teks yang mendalam hingga evaluasi sentimen masyarakat global, dari memetakan jaringan diplomasi hingga menerapkan teknik analisis *set by set*, buku ini menggali berbagai aspek metode komputasi dalam konteks hubungan internasional.

Buku ini dibagi dalam 5 bab, dengan bab pertama membahas mengenai mengapa metode komputasi penting dalam konteks analisis hubungan internasional. Pengenalan volume data yang eksplosif, yang terus meningkat berkat kemajuan teknologi, membutuhkan pendekatan yang lebih efisien dan akurat. Metode komputasi memungkinkan pengolahan data yang cepat dan akurat, membantu para analis untuk mengeksplorasi pola-pola yang tersembunyi di tengah data besar. Bab ini juga menyajikan data mengenai sejumlah riset yang mempergunakan metode komputasi dalam menganalisis fenomena hubungan internasional.

Bab kedua menyoroti peran penting *text mining* dalam analisis hubungan internasional. Dengan menggunakan teknik-teknik linguistik komputasional, pembaca akan memahami bagaimana teks-teks diplomatik, pidato kebijakan, dan sumber-sumber lainnya dapat dianalisis secara mendalam untuk mengungkapkan tren, kecenderungan, dan makna yang mendasar.

Bab ketiga membahas bagaimana analisis sentimen dapat diterapkan dalam konteks hubungan internasional. Metode komputasi memungkinkan analisis perasaan dan pandangan

masyarakat terhadap berbagai isu global. Dengan memantau respons sosial terhadap peristiwa dan kebijakan internasional, para analis dapat memahami reaksi publik dan potensi dampaknya pada dinamika internasional.

Bab keempat menjelaskan pentingnya analisis jaringan dalam studi hubungan internasional. Metode komputasi memungkinkan pembuatan peta dan analisis jaringan diplomatik, kerja sama lintas negara, dan interaksi internasional lainnya. Dengan memahami struktur jaringan ini, kita dapat mengidentifikasi aktor-aktor kunci, aliansi, dan potensi konflik dalam sistem internasional.

Bab terakhir mengilustrasikan penerapan metode komputasional melalui studi kasus yang spesifik dan berfokus pada analisis *set by set*. Teknik ini memungkinkan analisis yang lebih mendalam dengan memecah kompleksitas analisis menjadi *subset* yang terkelola. Studi kasus ini akan menggambarkan bagaimana metode komputasional dapat diaplikasikan dalam menganalisis situasi dunia nyata.

Dengan menggabungkan konsep teoretis dengan aplikasi praktis, buku ini akan membekali pembaca dengan wawasan dan keterampilan yang diperlukan untuk memahami dan menganalisis hubungan internasional dalam cara yang baru dan lebih terperinci. Metode komputasi menghadirkan alat yang kuat untuk menghadapi tantangan kompleksitas dunia global saat ini, dan buku ini akan membantu pembaca menguasai alat tersebut dengan baik.

Daftar Isi

Pengantar	v
Daftar Isi.....	viii
Daftar Grafik.....	xi
Daftar Gambar.....	xii
Daftar Tabel	xiii
Bab I Mengapa Metode Komputasi?	1
I.1. Latar Belakang Metode Komputasi dalam Ilmu Hubungan Internasional.....	3
I.2. <i>Big Data</i> dalam Metode Komputasi.....	8
I.3. Metode Komputasi dalam Riset Hubungan Internasional.....	18
I.3.1. <i>Big Data</i> bagi Riset Deskriptif.....	18
I.3.2. <i>Big Data</i> bagi Riset Eksperimentasi	23
I.3.3. <i>Big Data</i> bagi Riset Prediktif	28
I.4. Penutup	33
Referensi.....	34
Bab II <i>Text Mining</i> dalam Penelitian Hubungan Internasional.....	39
II.1. Definisi <i>Text Mining</i>	41
II.2. Kumpulan Dokumen dalam Metode <i>Text Mining</i>	42
II.3. Pra-Pemrosesan dalam Metode <i>Text Mining</i>	45
II.4. <i>Topic Modeling</i>	50

II.5. Penerapan <i>Text Mining</i> menggunakan <i>Topic Modeling</i> dalam Riset Hubungan Internasional	53
II.6. Penutup	57
Referensi.....	58
Bab III Analisis Sentimen	60
III.1. Metode dan Teknik Analisis Sentimen.....	62
III.1.1. Pendekatan Machine Learning	62
III.1.2. Pendekatan Rule Based.....	63
III.1.3. Analisis Semantik	65
III.1.4. Pendekatan Berbasis Leksikon.....	66
III.2. Contoh Analisis Sentimen secara Praktis (Menggunakan Leksikon).....	68
III.3. Contoh Analisis Sentimen menggunakan Pembelajaran Mesin (<i>Supervised Learning</i>)	71
III.4. Visualisasi Hasil Sentimen Analisis.....	73
III.5. Penggunaan Analisis Sentimen dalam Riset Hubungan Internasional.....	75
III.6. Penutup	76
Referensi.....	78
Bab IV Analisis Jaringan dalam Ilmu HI	79
IV.1. Jaringan	80
IV.2. Representasi Jaringan.....	83
IV.3. Struktur Jaringan	86
IV.4. Aliran Jaringan.....	87
IV.5. Pengukuran dalam Jaringan	89
IV.5.1. Sentralitas.....	89
IV.5.2. Densitas	95
IV.5.3. Konektivitas.....	96
IV.5.4. Sentralisasi	96

IV.6. Contoh Penggunaan Analisis Jaringan dalam Riset HI.....	97
Referensi.....	99
Bab V Aplikasi Metode Komputasional <i>Set by Set</i>.....	101
V.1. Analisis Sentimen.....	102
V.2. Analisis Jaringan.....	106
Profil Penulis	112

deepublish / publisher

Daftar Grafik

Grafik I.1	Jumlah Publikasi <i>Computational Social Science</i> Berdasarkan Tahun.....	6
Grafik I.2.	Skor Sentimen untuk Pasangan Negara	21

Daftar Gambar

Gambar I.1.	7 V's Karakter Big Data	11
Gambar I.2.	Diagram Jaringan dari Sentimen Negatif dan Positif	22
Gambar I.3.	Akurasi dalam <i>Validation Sets</i>	26
Gambar I.4.	Akurasi dalam <i>Training Set</i>	27
Gambar I.5.	ER dalam <i>Validation Set</i>	27
Gambar II.1.	Proses Tokenisasi	46
Gambar II.2.	Proses Filterisasi	49
Gambar II.3.	Proses <i>Stemming</i>	50
Gambar III.1	Kata Positif	74
Gambar III.2.	Kata Negatif	74
Gambar IV.1.	Pemain Sepak Bola Profesional	81
Gambar IV.2.	Pemain Sepak Bola dengan Hubungan Tim Sepak Bola yang Sama	82
Gambar IV.3	Pemain Sepak Bola yang Pernah Bermain Bersama	82
Gambar IV.4.	Jaringan dengan Relasi Berarah	83
Gambar IV.5.	Terdapat Tiga Jaringan Bagian dengan di mana A Merupakan Singleton; B dan C Merupakan Dyad; D, E, dan F Merupakan Triad	86
Gambar IV.6.	Ilustrasi Aliran Jaringan	88
Gambar IV.7.	Perbandingan Densitas Dua Jaringan	95

Daftar Tabel

Tabel I.1.	Perhitungan dari Berbagai Metode.....	25
Tabel I.2.	Regresi Ekspor pada GNP, Populasi, Jarak, Aliansi, dan Perang, 1905-1985.....	31
Tabel I.3.	Regresi Ekspor pada GNP Per Kapita, Jarak, dan Aliansi (Kecuali Uni Soviet, 1905-1985).....	32
Tabel II.1.	Temuan Topik dengan Metode LDA.....	55
Tabel III.1.	Contoh Penetapan Skor Sentimen untuk "Makanannya Enak Sekali!"	70
Tabel III.2.	Contoh Penetapan Skor Sentimen untuk "Pelayanannya Buruk tapi Makanannya Enak Sekali"	71
Tabel III.3.	Pelabelan Manual	72
Tabel IV.1.	Representasi dalam Bentuk Daftar Ketetangaan.....	84
Tabel IV.2.	Daftar Ketetangaan dengan Bobot.....	84
Tabel IV.3.	Daftar Ketetangaan dari Jaringan Berarah pada Gambar IV.4.....	84
Tabel IV.4.	Matriks Ketetangaan dari Gambar IV.3.	85
Tabel IV.5.	Matriks Ketetangaan dari Jaringan Berarah pada Gambar IV.4.....	85

deepublish / publisher



Bab I

Mengapa Metode Komputasi?

Sebagai seorang peneliti, dosen, maupun analis, kejadian mana yang pernah anda alami:

1. Jadwal wawancara dengan narasumber yang sudah dibuat jauh-jauh hari tiba-tiba harus dibatalkan secara mendadak.
2. Proses pengambilan data membutuhkan waktu dan biaya yang banyak, terkadang untuk hal-hal yang sifatnya non teknis.

Satu hal yang sering tidak disadari oleh seorang ilmuwan sosial adalah, dalam dunia yang terdigitalisasi seperti sekarang ini, data terkumpulkan secara konstan dan dalam jumlah yang besar, menunggu untuk diambil dan dianalisis.

Seorang publik figur membuat *posting*-an di Twitter? Data.

Akun YouTube media *upload* video berita? Data.

Seorang diplomat memberikan pidato di hadapan Sidang Umum PBB? Data.

Media masa membuat artikel berita *online*? Data.

Dunia yang terdigitalisasi membuat riset dalam hubungan internasional tidak jarang melibatkan sejumlah data yang kompleks. Hal yang diteliti mungkin masih sama, fenomena internasional seperti perilaku negara, perubahan kebijakan negara, dampak konflik, perang dan damai dan lain sebagainya. Namun cara melakukannya semakin lama semakin berkembang. Ada

banyak teori yang bermunculan untuk menjelaskan fenomena dalam hubungan antar negara tersebut. Hal yang sama juga berlaku dalam hal metode; terus berkembang (Rahmadi 2011). Hubungan internasional merupakan bidang yang kaya akan kontestasi metodologis yang mengacu pada tradisi penelitian yang beragam. Meski demikian cara atau prosedur riset yang beragam tersebut tidak boleh membuat peneliti hubungan internasional gagal dalam menghasilkan standar penelitian yang berkualitas (Lamont 2015).

Seperti halnya dalam riset sosial lainnya, para peneliti studi hubungan internasional juga sering dihadapkan pada tantangan data yang besar dan kompleks, seperti survei sosial, data media sosial, dan data administratif. Riset mengenai pengaruh media sosial dalam gerakan sosial misalnya, menjadi salah satu isu hubungan internasional yang penting untuk dikembangkan mengingat luasnya jaringan dan pengguna media sosial. Riset dalam bidang ini tidak hanya penting namun juga menarik mengingat beberapa gerakan sosial berhasil membentuk identitas transnasional (Edrington and Lee 2018; Phalapong 2022) dan gerakan lain berujung dengan revolusi (Al-Rawi and Groshek 2018; Wolfsfeld, Segev, and Sheaffer 2013). Tumbuh pesatnya media digital juga menyediakan ruang bagi riset mengenai jejaring diplomasi digital dan pengaruh wacana yang berkembang dalam dunia maya (Bjola and Jiang 2015; Kent and Taylor 1998; Sevin and Ingenhoff 2018).

Data yang tersedia dalam jumlah yang besar dan bisa diakses dengan mudah dan murah mampu mengubah cara peneliti mencari dan menyusun data. Aktor hubungan internasional kini memanfaatkan jejaring *online*, media masa mulai tergeser sosial media, dan opini publik dapat dengan mudah dilihat dan bahkan dipengaruhi. Hal ini menjadi tantangan sekaligus kesempatan bagi peneliti. Data yang tersedia dalam jumlah yang besar di satu sisi

memberikan kemudahan, namun di sisi lain dapat membingungkan jika tidak tahu cara mengolahnya. Berbagai teknik baru kini tersedia untuk menganalisis kumpulan data yang besar dan kompleks ini. Dengan meminjam metode dari ilmu komputer dan pemrograman, ilmuwan HI bisa dengan mudah mendapatkan temuan dari teks yang jumlahnya besar, berubah dengan cepat, dan sekilas tidak teratur (Molina and Garip 2019; Nelson 2020).

Kebutuhan riset atas data yang besar (*big data*) akan menyulitkan peneliti untuk melakukan riset tanpa bantuan metode riset yang sesuai. Metode komputasi, seperti analisis statistik, analisis jaringan sosial, dan analisis teks, telah membuka peluang baru bagi periset sosial dalam mengolah dan menganalisis data. Dengan menggunakan algoritma komputasi yang canggih—namun sebenarnya mudah dioperasikan—para peneliti dapat mengidentifikasi korelasi, tren, dan pola yang tersembunyi dalam data, yang pada gilirannya memberikan pemahaman yang lebih mendalam tentang fenomena sosial yang sedang dipelajari. Metode komputasi ini memungkinkan peneliti untuk mengeksplorasi data secara komprehensif dan menyeluruh, membantu mereka mengungkap wawasan baru dan menganalisis hubungan yang rumit antara variabel-variabel yang ada. Dengan cara ini, metode komputasi menjadi sarana yang penting dalam menghadapi tantangan data yang kompleks dalam riset sosial, membuka pintu bagi kemajuan penelitian yang signifikan.

I.1. Latar Belakang Metode Komputasi dalam Ilmu Hubungan Internasional

Metode komputasi merupakan metode riset yang dipergunakan untuk mendapatkan jawaban dari data dalam jumlah yang sangat besar dengan memanfaatkan kekuatan komputasi dari mesin. Bayangkan anda dikirim *file* Excel yang berisi 10.000 *tweet* yang ditulis oleh Antonio Guterres, kemudian anda diminta untuk menghitung berapa kali kata "*peacekeeping*" muncul. Mungkin pada

baris ke 1000 atau 2000 mata anda sudah lelah dan kemudian menyerah. Namun jika anda meminta komputer melakukan itu, maka tugas itu akan selesai dalam 10-20 detik saja.

Mengacu Ünver, metode komputasi dalam studi hubungan internasional dapat menjadi jembatan yang menghubungkan kesenjangan metodologis karena metode ini menguji teori hubungan internasional melalui analisis data yang terorganisir dengan baik. Seorang peneliti hubungan internasional dapat membangun kemampuan teknisnya melalui metode ilmu *computer* dalam berbagai aplikasi seperti Python, R, QGIS, atau ArcGIS. Metode komputasional bahkan lebih daripada suatu pilihan metode dalam mazhab kuantitatif atau kualitatif—sebab metode ini memberikan dasar yang kuat untuk menggabungkan keduanya (Taninecz Miller *et al.* 2019; Ünver 2018). Metode komputasi dalam ilmu hubungan internasional yang juga disebut sebagai *Computational International Relations* oleh Ünver (2018), berasal dari *Computational Social Science*.

Melanie (1999) dan Ünver (2018) mencatat perkembangan penggunaan metode komputasi dalam ilmu sosial atau *Computational Social Science* telah dimulai semenjak tahun 1950-1960-an ketika model matematika untuk mewakili konsep “agensi” muncul untuk menjelaskan perilaku sosial dan ekonomi individu dalam sebuah lingkungan. Secara sederhana Edelman *et al.* (2020) mendeskripsikan *Computational Social Science* sebagai bidang interdisipliner yang memajukan teori perilaku manusia dengan menerapkan teknik komputasi untuk mengolah data dari situs media sosial, internet, atau arsip digital lainnya seperti catatan administrasi. Salah satu contoh permodelan berbasis agen dipergunakan oleh Salgado untuk menyimulasikan bagaimana interaksi sosial yang berlangsung dalam sebuah lingkungan di antara agen yang otonom dan heterogen guna menjelaskan fenomena sosial tertentu. Pada tahun 80-an metode komputasi

berkembang dalam bentuk *Data Mining*¹, *Genetic Algorithms*² dan *System Dynamics Models* dalam ilmu sosial.³ Penggunaan pemodelan dan sistem jaringan terus berkembang hingga tahun 90-an. Sedangkan tahun 2000-an, *big data* (data besar) mulai dipergunakan dalam riset mengenai pengambilan keputusan, pendekatan perilaku, interaksi sosial dan analisis sistemik.

Pada tahun 2012, Conte *et al.* mengidentifikasi tiga aliran utama dalam *Computational Social Science*, yaitu:

- Aliran Deduktif: Membangun teori makro dengan menggunakan model matematika.
- Aliran Generatif: Membangun teori mikro melalui pemodelan perilaku dan simulasi komputer.
- Aliran Kompleksitas Ilmu Pengetahuan: Menggunakan kumpulan data besar dan dinamis untuk menjelaskan dan memprediksi perilaku dan konsekuensinya.

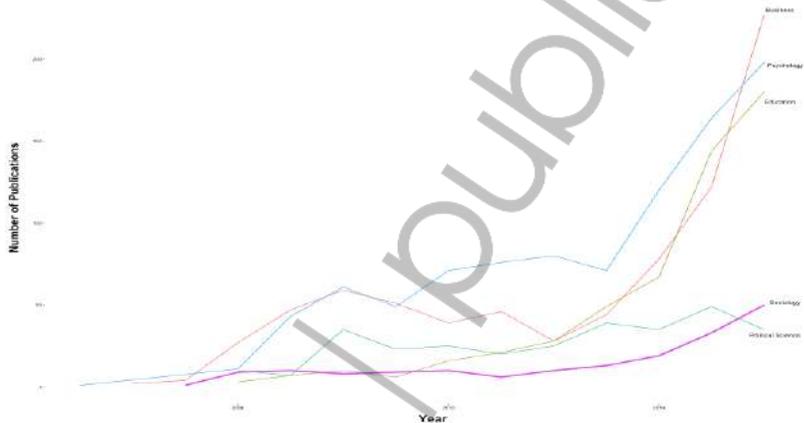
Dengan kemajuan *platform* digital dan media sosial, *Computational Social Science* kemudian berkembang menjadi ilmu sosial berbasis komputer yang memproses data sosial kuantitatif,

¹ Kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola dan hubungan dalam set data berukuran besar atau penggunaan program komputer untuk menyimulasikan perilaku manusia dalam populasi buatan, lihat Santosa Budi, 2007, *Data Mining: Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Yogyakarta: Graha Ilmu.

² Algoritma genetika adalah jenis algoritma optimisasi, yang digunakan untuk menemukan solusi optimal untuk masalah komputasi tertentu yang memaksimalkan atau meminimalkan fungsi tertentu. Algoritma genetika mewakili salah satu cabang bidang studi yang disebut komputasi evolusioner, di mana mereka meniru proses biologis reproduksi dan seleksi alam untuk memecahkan solusi yang 'paling cocok', lihat Melanie Mitchell. 1999. *An Introduction to Genetic Algorithms*. Cambridge: A Bradford Book the MIT Press.

³ Burns dalam risetnya mengilustrasikan peran ilmiah dimainkan oleh pendekatan sistem dalam ilmu sosial dan humaniora, melalui kontribusinya pada penggunaan bahasa yang sama, konseptualisasi, dan integrasi teoretis dalam menghadapi fragmentasi yang ekstrem (dan berkembang) di antara ilmu sosial (dan antara ilmu sosial dan ilmu alam), lihat Tom R. Burns. 2006. "The Sociology of Complex Systems: An Overview of Actor-System-Dynamics Theory". *World Futures*, 62

termasuk pengolahan *big data* internet secara *real-time*, menurut Ünver pada tahun 2018. Namun, meskipun perkembangannya pesat, metode komputasi dalam ilmu sosial masih lebih banyak digunakan dalam bidang ekonomi bisnis. Riset dengan metode komputasi dalam bidang politik masih relatif sedikit dibandingkan dengan ilmu sosial lainnya (Lihat Gambar 1.1).



Grafik I.1 Jumlah Publikasi *Computational Social Science* Berdasarkan Tahun
Sumber: Edelman

Computational International Relations merupakan pendekatan untuk memahami berbagai aspek hubungan internasional dengan bantuan komputer dan teknologi. Pendekatan ini membahas topik seperti kekuasaan, konflik/perdamaian, perilaku negara, norma/lembaga internasional, sistem/tatanan dunia, dan juga aktor non-negara seperti lembaga swadaya masyarakat, perusahaan multinasional, media, kelompok agama, diaspora, dan militan.

Pada tahun 2017, SAGE menerbitkan sebuah buku yang menggabungkan metode penelitian kualitatif dan kuantitatif dengan menggunakan alat-alat seperti Kecerdasan Buatan, Data Besar, Penambangan Data, dan Analisis Geospasial dalam studi

Computational International Relations. Buku ini merupakan terobosan dalam kajian hubungan internasional, karena sebelumnya media sosial hanya dianggap sebagai *platform* untuk bersosialisasi.

Pentingnya pendekatan ini semakin jelas pada tahun 2016 ketika data yang dihasilkan dari media sosial dan *platform* lainnya melampaui jumlah data yang pernah ada sebelumnya dalam riset akademis. Hal ini memberikan peluang baru untuk mengkaji hubungan internasional dengan memanfaatkan komputer dan rekayasa perangkat lunak untuk menganalisis data yang sangat besar ini. Namun, di sisi lain, hal tersebut juga menimbulkan pertanyaan tentang cara penggunaan metode ini secara operasional dan definisi yang tepat dalam studi ini (Hadi 2020).

Namun, sejalan dengan ilmu hubungan internasional yang masih tumbuh dan berkembang, maka metode riset komputasi dalam hubungan internasional pun masih akan terus tumbuh dan berkembang guna mendukung riset yang melibatkan *big data*. perkembangan ilmu hubungan internasional akan terus mendorong pertumbuhan dan perkembangan metode riset komputasi dalam bidang ini, terutama untuk mendukung analisis *big data*. Selain riset media sosial, pemetaan atau analisis geospasial juga memiliki peran penting dalam menganalisis isu-isu perbatasan dan kebijakan yang terkait dengan geografi.

Teknologi digital juga memberikan dukungan berharga bagi peneliti hubungan internasional dalam menganalisis perdagangan internasional melalui kemampuan kalkulasi digital. Selain itu, perkembangan teknologi *machine learning* memungkinkan riset naratif, di mana peneliti dapat mempelajari narasi yang ada di media sosial atau *platform online* lainnya. Hal ini memungkinkan analisis lebih mendalam dan pemanfaatan data secara lebih efisien untuk memahami berbagai isu dalam hubungan internasional. Dengan adanya kemajuan ini, metode komputasi akan terus menjadi alat yang sangat berharga bagi peneliti hubungan

internasional untuk menjelajahi dan menganalisis berbagai aspek kompleks dari dinamika internasional.

I.2. *Big Data* dalam Metode Komputasi

Metode komputasi digunakan oleh peneliti sosial untuk mengatasi tantangan riset yang melibatkan data dalam jumlah besar yang sulit dicari, diklasifikasi, dan dianalisis tanpa bantuan teknologi komputer. *Big data* secara umum mengacu pada data yang memiliki volume besar, berkecepatan tinggi, kompleksitas tinggi, dan memerlukan teknologi canggih untuk mengolah dan menganalisis informasinya.

Menurut TechAmerica Foundation's Federal Big Data Commission (2012), *big data* adalah data dengan volume besar yang memerlukan teknik dan teknologi khusus untuk diproses. International Data Corporation (IDC) mendefinisikan *big data* sebagai teknologi dan arsitektur baru yang memungkinkan ekstraksi nilai ekonomis dari data dalam volume besar dengan kecepatan tinggi. IBM menyatakan bahwa *big data* adalah kumpulan data dengan volume tinggi, kecepatan tinggi, dan variasi tinggi yang melampaui kapasitas basis data relasional tradisional. Sumber data semakin kompleks karena dipengaruhi oleh kecerdasan buatan (AI), perangkat seluler, media sosial, dan *Internet of Things* (IoT) (IBM n.d.).

Meskipun ada definisi yang beragam, *big data* memiliki peran yang signifikan dalam riset. Dalam sebuah survei *online* pada tahun 2012 oleh Harris Interactive terhadap 154 eksekutif Global Suites, ditemukan bahwa setiap narasumber memberikan pemahaman yang berbeda tentang apa yang dimaksud dengan *big data*, ada yang fokus pada definisi dan ada pula yang menekankan pada fungsi *big data*. (Gandomi and Haider 2015).

Gandomi dan Haider (2015) mendefinisikan *big data* berdasarkan karakteristiknya yang disebut "3 V's", yang sejalan

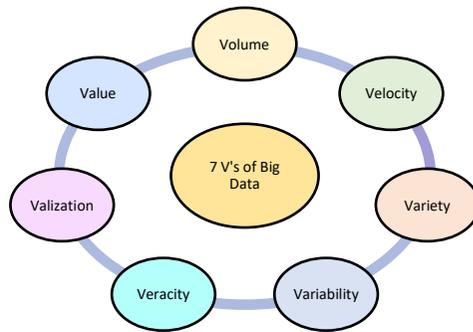
dengan definisi yang dibuat oleh IBM. Ketiga V's tersebut adalah *Volume*, *Variety*, dan *Velocity*.

- *Volume*: Karakteristik *volume* menunjukkan bahwa *big data* memiliki ukuran yang sangat besar, biasanya dihitung dalam terabyte atau bahkan petabyte. Sebagai perbandingan, satu terabyte bisa menyimpan data setara dengan 1500 CD atau 220 DVD, dan satu petabyte setara dengan 1024 terabyte. Definisi volume data besar bersifat relatif dan bervariasi tergantung pada waktu dan jenis data. Apa yang dianggap sebagai data besar hari ini mungkin tidak lagi dianggap demikian di masa depan karena kemajuan kapasitas penyimpanan data.
- *Variety*: Karakteristik *variety* menunjukkan bahwa *big data* mencakup beragam jenis data, seperti data tabulasi dan data video. Dua kumpulan data dengan ukuran yang sama mungkin memerlukan teknologi manajemen data yang berbeda karena perbedaan jenis dan bentuk data. Oleh karena itu, definisi *big data* juga bergantung pada jenis atau bentuk data yang sedang dianalisis.
- *Velocity*: Karakteristik *velocity* menunjukkan bahwa *big data* dihasilkan atau berubah dengan kecepatan tinggi. Data bisa berasal dari sumber yang berbeda secara *real-time* atau diperbaharui dengan frekuensi tinggi. Proses pengumpulan, penyimpanan, dan analisis data *big data* harus mampu menangani kecepatan ini.

Karena karakteristik yang kompleks dan bervariasi, tidak ada ambang batas yang tetap untuk definisi volume data besar. Pengertian *big data* dapat berubah seiring dengan perkembangan teknologi dan jenis data yang diproses. Hal ini menunjukkan bahwa definisi *big data* adalah relatif dan selalu berada dalam konteks tertentu. Kedua, *Variety* yang merujuk pada heterogenitas struktural dalam kumpulan data.

Seperti Gandomi dan Haider (2015), Zafar *et al.* (2021) juga menggambarkan *big data* menggunakan 7V's. Selain *Volume*, *Velocity*, dan *Variety*, Zafar *et al.* menambahkan empat karakteristik lainnya, yaitu:

- **Variabilitas:** Merupakan proses pembentukan interpretasi dan pemahaman yang benar terhadap data mentah dengan makna yang akurat. Dalam *big data*, data dapat sangat bervariasi dan kompleks, sehingga penting untuk mengelola variasi ini dengan tepat agar dapat mengambil *insight* yang benar dan relevan.
- **Veracity:** Mengacu pada keakuratan dan keandalan data yang dikumpulkan. *Big data* seringkali mencakup data dari berbagai sumber, dan *veracity* berfokus pada memastikan bahwa data yang digunakan dalam analisis adalah akurat dan dapat dipercaya, serta menyisihkan data yang tidak berkualitas atau kurang bermutu dari sistem.
- **Visualisasi:** Berkaitan dengan cara menyajikan data kepada pengguna untuk pengambilan keputusan. Visualisasi yang efektif dapat membantu menggambarkan pola dan informasi yang tersembunyi dalam *big data* dengan lebih jelas dan mudah dipahami.
- **Value (Nilai):** Karakteristik ini menekankan pentingnya menambahkan nilai pada data besar sehingga dapat memberikan manfaat dan membantu mengurangi biaya. Dengan menganalisis data secara tepat, *big data* dapat memberikan wawasan yang bernilai bagi organisasi atau perusahaan.



Gambar I.1. 7 V's Karakter Big Data
Sumber: (Zafar)

Big data memberikan banyak manfaat, seperti menyimpan dan mengolah data dalam volume besar dengan akurasi yang tinggi (Almeida 2017; Ijaz Baig, Shuib, and Yadegaridehkordi 2019). Dengan *big data*, kita bisa mengatur operasi dalam berbagai bidang dengan lebih efisien. Selain itu, *big data* memberikan kemampuan untuk melihat keragaman dalam masyarakat dengan jelas dan cepat melalui analisis menggunakan teknologi *machine learning*.

Big data juga mengatasi masalah dalam penelitian sosial di mana subjek penelitian seringkali mengubah perilakunya saat tahu sedang diobservasi. Data-data yang sering digunakan dalam analisis *big data* adalah data yang muncul secara alamiah. Sebagai contoh, data *twit* yang dibuat oleh *user* merupakan murni hasil keinginan *user*, tanpa ada perasaan diawasi. Ini membuat data lebih natural dan representatif (Salganik 2018).

Keunggulan lain selain dapat mengumpulkan dan mengolah data dalam jumlah yang sangat besar adalah harganya yang murah. Sejumlah perusahaan telah menghasilkan teknologi penyimpanan data dan menawarkan mesin pemrosesan data dengan harga yang semakin hari semakin murah namun dengan kualitas yang semakin baik (Almeida 2017). *Big data* juga mampu membantu peneliti untuk menghindari kesimpulan yang diambil

dari terlalu sedikitnya data. Data yang besar memastikan peneliti memperoleh regularitas dari setiap peristiwa yang menjadi kajiannya (Salganik 2018) Melalui *big data*, maka persoalan “representasi” yang seringkali menjadi persoalan dalam riset sosial karena mendasarkan diri pada penggunaan sampel acak dapat dihindarkan.

Big data juga dapat membantu peneliti membuat prediksi dengan lebih baik. Cole menunjukkan bagaimana kekuatan komputasi komputer dapat digunakan untuk memprediksi penyebaran penyakit menular. Berdasarkan data mengenai pola penyebaran sebelumnya, mesin dapat memprediksi wilayah mana saja yang akan segera terjangkiti dengan mempertimbangkan karakter penyakit, mobilitas penderita, suhu, curah hujan dan lain sebagainya. Dengan prediksi yang cepat dan akurat, langkah pencegahan dapat dilakukan dengan lebih efektif. Di bidang kesehatan dan ekonomi, dokter dan ahli ekonomi bisa menemukan informasi penting dari data yang mereka kumpulkan. Misalnya, dokter dapat menemukan obat-obatan baru untuk penyakit berdasarkan data pasien yang terkumpul. Sedangkan di bidang ekonomi, data konsumsi dapat membantu memprediksi jenis barang apa yang akan diminati oleh pasar.

Selain itu, penggunaan data besar (*big data*) juga bisa menciptakan produk dan layanan baru yang dibutuhkan oleh masyarakat. Hal ini berlaku di berbagai bidang seperti ekonomi, astronomi, kesehatan, dan telekomunikasi. Semua kemajuan ini telah menciptakan disiplin ilmu baru yang disebut sains data (*data science*), yang menggabungkan pengetahuan dari berbagai mata pelajaran seperti psikologi, statistik, ekonomi, ilmu sosial, ilmu jaringan, dan ilmu kompu (Bhadani and Jothimani 2016).

Perkembangan teknologi komputasi, khususnya prosesor dan memori yang semakin besar kapasitas dan kecepatannya telah memainkan peran krusial dalam meningkatkan kemampuan *big*

data untuk mengakses data dalam volume besar secara waktu nyata (*real time*). Dulu, mengakses dan menganalisis data besar membutuhkan waktu yang lama karena perlu mentransfer data dari penyimpanan ke memori sebelum dapat diproses. Namun, dengan teknologi pemrosesan data dalam memori, data besar dapat diakses langsung dari memori perangkat keras tanpa perlu mengalami keterlambatan karena transfer data.

Ini berarti bahwa analisis data besar dapat dilakukan secara cepat dan efisien, memungkinkan pengambilan keputusan yang lebih tepat waktu dan mendukung berbagai aplikasi yang memerlukan akses data secara instan. Kemampuan untuk mengakses data bervolume besar secara waktu nyata ini telah membuka peluang baru dalam berbagai bidang, seperti penelitian ilmiah, analisis pasar, prediksi cuaca, pengelolaan lalu lintas, dan banyak lagi.

Selain itu, perkembangan ini juga berdampak pada kemampuan *big data* dalam mendeteksi penipuan digital. Penipuan digital mengacu pada usaha membuat representasi palsu atau menyalahgunakan posisi untuk mendapatkan keuntungan finansial atau manfaat tidak semestinya, seperti penipuan asuransi, kartu kredit, pembobolan akun, dan penipuan telekomunikasi (Abbassi, El Alaoui, and Gahi 2022).

Mendeteksi penipuan bukanlah tugas yang mudah karena memerlukan informasi dan pengetahuan mendalam tentang berbagai sifat penipuan, seperti intrusi jaringan, transaksi dan pengadaan, penyalahgunaan komputer, pemalsuan dan manipulasi data, klaim palsu, dan lain sebagainya. Namun, dengan mempelajari konteks di mana penipuan terjadi, kita dapat menetapkan hubungan antara anomali dan insiden penipuan. Contohnya, kenaikan harga suatu komoditas langka yang permintaannya tinggi, adanya transaksi berkali-kali dalam waktu singkat, peningkatan impor suatu komoditas yang permintaannya

tidak berubah, dan adanya transaksi dalam periode waktu yang tidak teratur seperti transaksi kartu kredit pada waktu tertentu dapat menjadi petunjuk adanya potensi penipuan (Mojtahed 2019).

Dengan menggunakan *big data* dan teknologi analisis canggih, sistem dapat mengidentifikasi pola-pola mencurigakan seperti di atas dan memberikan peringatan dini atas potensi penipuan, sehingga membantu melindungi penyedia layanan atau institusi dari risiko penipuan dan mencegah kerugian finansial yang tidak perlu (Balachandran and Prasad 2017).

Ilustrasi di atas menunjukkan bahwa analisis *big data* bisa menjadi alat yang sangat efektif. Analisis *big data* dilakukan dengan menggunakan berbagai teknik, seperti *machine learning*, *data mining*, dan *deep learning*. Sudah ada banyak perangkat lunak yang digunakan untuk melakukan analisis ini.

Dengan bantuan perangkat lunak tersebut, organisasi dan institusi bisa memproses dan menganalisis data besar dengan lebih cepat dan efisien. Dengan cara ini, mereka dapat menemukan pola-pola penting dalam data, melihat kecenderungan yang sedang terjadi, dan membuat keputusan yang lebih cerdas berdasarkan informasi dari data besar.

Analisis *big data* memiliki peran yang penting dalam berbagai bidang, seperti bisnis, ilmu pengetahuan, kesehatan, dan lainnya. Dengan kemampuannya dalam mengolah data besar, analisis *big data* membantu organisasi dalam menghadapi tantangan dan memanfaatkan peluang dengan lebih baik.

Namun demikian, *big data* tidak luput kelemahan. Kelemahan tersebut antara lain kualitas kemampuan komputasi komputer (Almeida 2017). Tidak semua perangkat komputer mampu mengoperasikan program tertentu, karena keterbatasan spesifikasi.

Kelemahan lain adalah kualitas data yang ada. Data yang selalu berubah walau di satu sisi memberikan kuantitas data yang

banyak, namun di sisi lain kualitas data yang terkumpul juga bisa mengalami penyimpangan. Secara khusus, data dapat berubah dalam tiga cara yaitu penyimpangan populasi, perilaku dan penyimpangan sistem (Salganik 2018). Penyimpangan populasi merupakan penyimpangan yang terjadi ketika pengguna media sosial, blog atau sistem digital lainnya berubah. Misalnya pengguna Twitter dalam masa kampanye presiden dalam kurun waktu tertentu. Pada kurun waktu kampanye, kelompok tertentu sangat sering mempergunakan Twitter agar suara mereka memengaruhi pemilik, namun pada masa kampanye telah usai mereka bukan lagi pengguna yang aktif. Riset (Kruikemeier 2014) menemukan bahwa penggunaan Twitter di masa kampanye presiden meningkat dan kandidat yang memanfaatkan Twitter lebih banyaklah yang memenangkan suara pemilih. Hal ini dapat mempengaruhi data yang terkumpul, sebab data yang dikumpulkan dalam masa pemilu mestinya dianalisis dan diinterpretasi secara berbeda, dengan mempertimbangkan konteks pemilu. Peneliti yang tidak teliti dapat menarik kesimpulan yang salah jika tidak dengan saksama mempertimbangkan konteks ini.

Kedua, penyimpangan perilaku yaitu perubahan dalam cara orang menggunakan media sosial atau media digital lainnya. Misalnya selama protes terhadap kudeta yang terjadi di Myanmar, pengguna Twitter mempergunakan sejumlah tagar tertentu untuk menarik perhatian pengguna dan membangun identitas dan persepsi yang sama terhadap peristiwa politik yang terjadi di Myanmar (Temby 2021; Utama 2022). Melalui identitas dan persepsi yang sama tersebut maka, lebih mudah massa digiring pada perilaku sosial tertentu. Hal ini juga dapat mempengaruhi data yang terkumpul sebab ada penggiringan opini. Ketiga, penyimpangan sistem yaitu perubahan dalam sistem itu sendiri. Misalnya, ketika Twitter mengubah jumlah karakter dalam *twit*-nya maka hal ini juga akan memengaruhi perilaku pengguna dan

kontennya seperti halnya ketika Instagram ataupun TikTok mengubah, menambah atau mengurangi fitur yang dimilikinya.

Faktor akses data juga menjadi tantangan tersendiri bagi peneliti, karena meskipun data tersedia di dalam website, tidak semua pemilik website mengizinkan setiap orang mengakses data tersebut tanpa izin. Dalam konteks yang sama, akses terhadap data pribadi peneliti juga menjadi persoalan dalam riset *big data* (Balachandran and Prasad 2017; Bhadani and Jothimani 2016; Salganik 2018). *Big data* tidak mampu menembus sejumlah halangan lain untuk dapat menyediakan data dengan lengkap. *Big data*, atau data dalam skala besar, seringkali dianggap lengkap. Namun, kenyataannya tidak selalu demikian.

Mirip dengan survei yang sering dilakukan oleh ilmuwan sosial, *big data* juga bisa memiliki kekurangan. Menurut Salganik, *big data* biasanya kurang tiga jenis informasi penting. Pertama, data demografis, yang memberikan informasi tentang latar belakang subjek penelitian. Kedua, data tentang perilaku subjek di *platform* lain di luar yang sedang diteliti. Ketiga, data yang dibutuhkan untuk mengaplikasikan atau menguji teori tertentu. Kekurangan data ini bisa disebabkan oleh berbagai alasan. Bisa jadi karena akses ke data yang sengaja ditutup atau tidak disediakan oleh pemilik data. Namun, alasan lain juga bisa karena adanya hambatan hukum, bisnis, atau etika yang mencegah kita mengakses data tersebut.

Perusahaan biasanya memanfaatkan *big data* untuk mengumpulkan banyak informasi, misalnya tentang pelanggan atau calon pelanggan mereka. Tetapi hal ini menimbulkan pertanyaan, siapa sebenarnya yang memiliki hak atas data tersebut? Setiap institusi yang memiliki data tentang pegawai atau pengguna layanannya tentu akan berusaha untuk melindungi data tersebut. Mereka perlu menjaga agar data tersebut tidak disalahgunakan dan untuk menjaga kepercayaan serta kenyamanan

dari pegawai atau pengguna layanan mereka (Cole, Nelson, and McDaniel, 2015).

Salah satu keterbatasan dari *big data* adalah peneliti seringkali lupa bahwa data yang tersedia bukanlah hasil alami. Ini karena sistem digital secara sengaja dirancang untuk mendorong perilaku tertentu. Proses ini dalam membentuk data agar sesuai dengan pola tertentu dikenal sebagai *Algorithmic Confounding* (Salganik, 2018). Ambil contoh Facebook. Platform ini sering merekomendasikan untuk memperluas jaringan pertemanan berdasarkan teman yang sudah ada, bukan menyarankan individu yang sama sekali tidak memiliki koneksi bersama. Rekomendasi ini tidak dibuat berdasarkan atribut seperti jenis kelamin, hobi, atau pekerjaan. Oleh karena itu, data yang dikumpulkan tidak secara akurat mewakili jaringan sosial *online* seseorang.

Penelitian oleh Chaney, Stewart, dan Engelhardt (2018) menunjukkan bahwa *Algorithmic Confounding* dapat memperkuat homogenitas dalam perilaku pengguna. Temuan ini menyarankan bahwa para perancang sistem tidak hanya harus mempertimbangkan bagaimana sistem mereka mempengaruhi perilaku pengguna, tetapi juga dampak potensial dari *Algorithmic Confounding*. Kelemahan lain yang ditandai oleh (Salganik 2018) adalah data yang dimasukkan ke dalam sistem seringkali tidak benar atau juga disebut dengan data kotor (*dirty data/noisy data*) karena tidak sesuai dengan tindakan nyata. Ridzuan menyebut data kotor sebagai data yang tidak akurat, tidak konsisten, dan tidak lengkap karena kesalahan yang ditemukan dalam kumpulan data. Dengan demikian, sebelum menerapkan penilaian terhadap sistem, peneliti harus memastikan kualitas informasi karena data yang kotor dan tidak seragam tidak dapat menghasilkan simpulan yang valid (Haughton *et al.* 2003). Oleh sebab itu, proses *cleaning* dan *pre-processing* dalam analisis data merupakan proses penting yang perlu dilakukan oleh data analis.

I.3. Metode Komputasi dalam Riset Hubungan Internasional

Riset dalam ilmu hubungan internasional dengan metode komputasi memang belum banyak dilakukan. Dengan demikian, ruang untuk mengembangkan riset ilmu hubungan internasional melalui metode komputasi masih terbuka cukup luas. Mempergunakan 3 strategi Salganik (2018) yang dapat dipergunakan oleh peneliti dalam memanfaatkan *big data*, riset yang dapat dikembangkan antara lain adalah riset yang bertujuan untuk menghitung/menjelaskan, memprediksi, dan eksperimentasi. Berikut adalah riset dalam studi hubungan internasional dengan mempergunakan sejumlah metode komputasi.

I.3.1. Big Data bagi Riset Deskriptif

Pembuatan keputusan luar negeri bukanlah hal yang mudah karena melibatkan jaringan aktor dan informasi yang sangat kompleks. Pembuatan kebijakan oleh seorang kepala negara tidak terlepas dari beberapa entitas seperti penasehat, lembaga, think-tank, NGO, perusahaan yang berpengaruh dan lain sebagainya. Namun, semakin banyak variabel yang digunakan dalam riset pengambilan keputusan maka semakin sulit untuk tidak hanya memodelkan antar-hubungan individu tetapi wawasan yang diperoleh justru menurun. Contoh penggunaan metode komputasi untuk menjelaskan atau mengukur dalam buku ini mengambil riset Shukla and Unger (2022) dengan judul *Sentiment Analysis of International Relations with Artificial Intelligence* dalam *Athens Journal of Sciences* 2022.

Shukla and Unger (2022) berpendapat bahwa informasi penting yang dapat digunakan dalam pembuatan keputusan adalah informasi mengenai hubungan antarnegara. Selain karena memang hal ini berdampak pada politik luar negeri, namun juga karena hubungan antarnegara dapat dipetakan melalui analisis

komputasional. Struktur politik dunia kini tidak hanya berfokus pada negara-negara saja. Ada organisasi besar lainnya, seperti PBB, WTO, dan NATO, yang juga mempengaruhi kebijakan global. Jadi, ketika kita mempelajari politik suatu negara, kita juga harus mempertimbangkan pengaruh dari organisasi-organisasi besar ini dan bagaimana mereka berinteraksi dengan negara tersebut.

Guna melakukan pemodelan, Shukla and Unger (2022) membangun jaringan berdasarkan pemberitaan yang dibuat di media dan mengukur sentimennya. Shukla and Unger (2022) mengumpulkan artikel berita yang berisi pemberitaan mengenai hubungan dua negara, kemudian mengukur sentimennya. Nilai rata-rata untuk masing-masing hubungan negara diukur. Kemudian hasilnya dibuat menjadi jaringan. Berita global yang menjadi sumber data diperoleh dari newsapi.org. Oleh karena keterbatasan akses berita, riset dibatasi untuk 100 pertanyaan berita per hari. Dengan memfokuskan pada 20 negara paling aktif secara geopolitik, Shukla menyusun daftar negara G20 yang meliputi: AS, Rusia, Cina, Inggris, Ukraina, Jerman, India, Iran, Israel, Turki, Korea Utara, Korea Selatan, Jepang, Australia, Arab Saudi, Prancis, Italia, Yunani, Pakistan, dan Indonesia. Jangka waktu penelitian dari 24 Maret 2021 hingga 23 Mei 2021 berhasil mengumpulkan 7.000 artikel berita yang langsung membahas hubungan antar dua negara. Selain menganalisis judul berita, riset juga menganalisis deskripsi berita. Hal ini memberikan skor sentimen yang jauh lebih akurat karena lebih banyak kata yang dapat dianalisis. Skor sentimen berkisar dari 0 hingga 1. Analisis sentimen dilakukan dengan metode BERT⁴.

⁴ Bidirectional Encoder Representations from Transformers atau BERT adalah sebuah model *machine learning* untuk NLP (*Natural Language Processing*). Bahasa secara historis sulit untuk 'dipahami' oleh komputer, meskipun komputer dapat mengumpulkan, menyimpan, dan membaca input teks tetapi tidak memiliki konteks bahasa dasar. Maka NLP membantu komputer untuk membaca, menganalisis, menafsirkan, dan memperoleh makna dari teks dan

Perlu diperhatikan bahwa skor sentimen yang diperoleh tidak mencerminkan tingkat intensitas seberapa baik atau buruk suatu berita, tetapi hanya menunjukkan apakah berita tersebut memiliki sentimen positif atau negatif. Dalam hal ini, tingkat keintensitasan suatu berita dapat disimpulkan dengan mengasumsikan bahwa peristiwa yang intens biasanya diungkapkan dengan jelas dan tegas, sedangkan peristiwa atau situasi yang tidak pasti cenderung mengarah pada pernyataan dan situasi berita yang tidak jelas. Oleh karena itu, skor sentimen yang diperoleh dapat diinterpretasikan sebagai indikator intensitas situasi positif atau negatif, bukan hanya sebagai probabilitas keakuratan.

Riset kemudian mulai dilakukan dengan melibatkan pemfilteran berita yang terkait dengan pasangan negara tertentu, misalnya semua berita pada hari tertentu antara AS dan Rusia. Kemudian berita ini di-*scraping* dan dianalisis dengan metode BERT untuk mengetahui sentimen yang terkait dengan berita tersebut. Jika ada banyak berita dalam satu hari, maka skor sentimen rata-rata dapat dipergunakan (mengambil waktu 30 hari terakhir untuk masing-masing pasangan negara). Nilai ini kemudian dipetakan melalui fungsi "tanh"⁵ untuk memperoleh nilai rata-rata sentimen analisis setiap pasangan negara. Skor sentimen yang diperoleh untuk setiap pasangan negara dapat dilihat sebagai **Gambar I.3**.

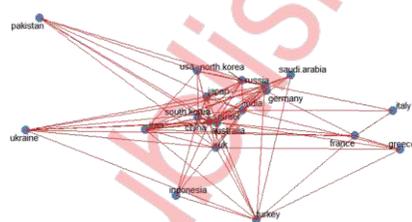
kata-kata yang diucapkan. Kinerja NLP menggabungkan linguistik, statistik, dan *machine learning* untuk membantu komputer dalam 'memahami' bahasa manusia.

⁵ Fungsi aktivasi tanh (*hyperbolic tangent*) adalah ungsi matematika mengubah input neuron menjadi angka antara -1 dan 1. Fungsi tanh sering digunakan dalam lapisan tersembunyi jaringan saraf karena memperkenalkan non-linearitas ke dalam jaringan dan dapat menangkap perubahan kecil pada masukan.

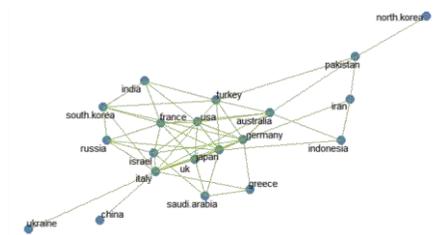


Grafik I.2. Skor Sentimen untuk Pasangan Negara
 Sumber: Shukla and Unger (2022)

Gambar I.3. menunjukkan rata-rata pergerakan skor sentimen selama 30 hari antara beberapa pasangan negara yang dipilih secara acak. Dari gambar tersebut, dapat melihat bahwa ada beberapa pasangan negara, seperti Rusia/Ukraina, Iran/Israel, dan AS/Cina, sistem berhasil menangkap sentimen yang tepat dan mencerminkan kenyataan. Namun, untuk pasangan negara lain, seperti AS/Jerman dan Rusia/Arab Saudi, sentimen cenderung tidak pasti atau tidak dapat sepenuhnya diandalkan. Alasannya adalah karena perkembangan skor sentimen cenderung linier. Hal ini menunjukkan bahwa jumlah berita yang dianalisis terbatas, sehingga skor sentimen dapat lebih rentan terhadap penafsiran yang salah. Meskipun demikian, sebagian besar perkembangan tetap mencerminkan kenyataan dalam hal menangkap sentimen positif atau negatif. Penting untuk dicatat bahwa skor sentimen tidak mencerminkan kondisi aktual dari sentimen antara dua negara, melainkan lebih sebagai probabilitas atau ketidakpastian bahwa sentimen tersebut positif atau negatif. Gambar I.4. berikut menunjukkan diagram jaringan baik dari sentimen negatif maupun positif sejumlah pasangan negara dalam 30 hari.



Network Diagram of Negative Sentiment Relationships from 04/23/2021-05/23/2021



Network Diagram of Positive Sentiment Relationships from 04/23/2021-05/23/2021

Gambar I.2. Diagram Jaringan dari Sentimen Negatif dan Positif
Sumber: Shukla and Unger (2022)

Diagram I.4. menunjukkan hubungan positif dan negatif antara negara-negara selama periode pengujian dari 04/23/2021 hingga 05/23/2021. Negara-negara yang berdekatan memiliki sentimen yang lebih baik satu sama lain dibandingkan dengan pasangan negara yang jauh. Artinya, negara-negara tetangga memiliki hubungan yang lebih baik daripada negara-negara yang berjauhan. Diagram jaringan ini membantu kita memahami jarak relatif dan hubungan sentimen internasional antara negara-negara. Shukla, melalui bantuan NLP dapat menjelaskan bagaimana hubungan antar negara melalui analisis berita dan menghasilkan temuan bahwa selama periode waktu yang dianalisis, hubungan antara pasangan negara seperti Pakistan/Rusia, Pakistan/AS, atau Turki/Yunani menjadi lebih buruk dibandingkan dengan hubungan antara pasangan negara seperti AS/Rusia, Jepang/Rusia, atau China/AS. Di sisi lain, hubungan antara pasangan negara seperti Prancis/AS, AS/Australia, atau AS/Jepang lebih baik dibandingkan dengan hubungan antara pasangan negara seperti Ukraina/Italia, Jepang/Indonesia, atau Turki/Pakistan.

I.3.2. Big Data bagi Riset Eksperimentasi

Guna menjelaskan bagaimana metode komputasi membantu dalam riset eksperimentasi, buku ini meminjam riset Schrodts (1991) yang berjudul *Prediction of Interstate Conflict Outcomes Using a Neural Network* dimuat dalam *Social Sciences Computer Review* pada tahun 1991. Schrodts (1991) menggunakan *neural networks* untuk memprediksi hasil konflik internasional, dan membandingkan keakuratannya dengan model yang dibangun menggunakan beberapa model pembanding. Alasan teoretis untuk mempergunakan jaringan saraf karena ia memiliki struktur yang dapat memprediksi perilaku politik (manusia). Sejalan dengan hal tersebut, maka jaringan saraf juga dapat dipergunakan untuk

memprediksi perilaku negara dalam hubungan internasional karena organisasi yang terstruktur sama dengan jaringan saraf dalam tubuh manusia. Misalnya sebuah kementerian luar negeri memperoleh input dari berbagai informasi yang diberikan oleh data intelijen. Semua laporan tersebut akan masuk pada manajemen menengah yang akan menimbang informasi yang dilaporkan secara positif, negatif, atau mengabaikannya sama sekali. Jika tertarik dan menganggap informasi sebuah isu penting, maka ia akan merekomendasikan informasi tersebut pada pembuat kebijakan. Jaringan saraf hanyalah analogi: manajer menengah dan atas jelas melakukan lebih dari sekadar menjumlahkan rekomendasi dan memasukkannya ke dalam fungsi nonlinier, sama seperti otak manusia yang mengandung komponen selain neuron dan sinapsis. Namun, kedua struktur itu secara struktural serupa dan dapat memastikan bahwa struktur organisasi pengambilan keputusan kebijakan luar negeri lebih menyerupai jaringan saraf daripada persamaan regresi. Sejauh struktur organisasi seperti itu menghasilkan keteraturan dalam perilaku internasional, keteraturan tersebut dapat dimodelkan asalkan variabel input yang sesuai teridentifikasi.

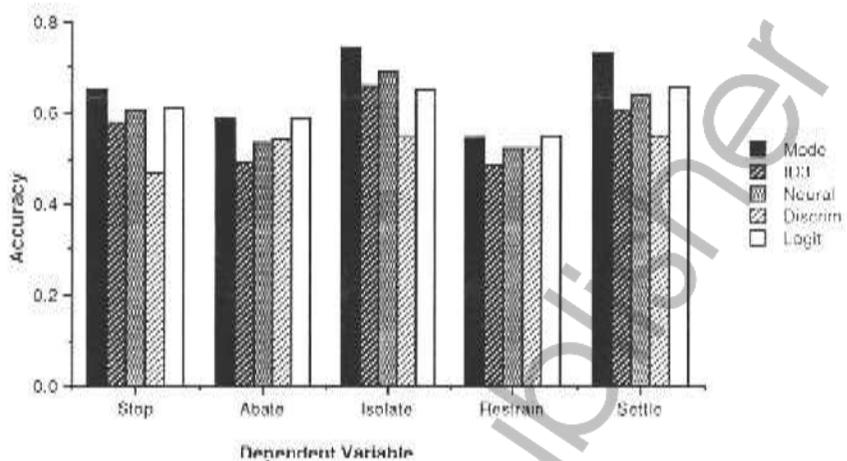
Schrodt (1991) mempergunakan data dari Butterworth's yang merupakan data mengenai konflik/perang yang juga menjelaskan tujuan dan tuntutan politik- kekuasaan tertentu yang memiliki dampak langsung pada perilaku nasional. Kumpulan data mencakup 310 kasus konflik antar negara dan 47 variabel yang terkait dengan karakteristik konflik, tindakan untuk mengelola konflik, dan hasil manajemen konflik. Jaringan saraf akan diuji dalam dua konfigurasi: pertama dengan menggunakan input interval, kedua menggunakan input biner. Data Butterworth menunjukkan manajemen konflik memberikan 5 hasil, yaitu: menghentikan permusuhan, meredakan konflik, mengisolasi konflik, menahan konflik, dan menyelesaikan konflik. Hasil Ini

diperlakukan sebagai variabel dependen, *output* dari saraf jaringan. Semua memiliki kode yang sama: hasil dari konflik sehubungan dengan perilaku akan menjadi: 0 = Tidak berbeda; 1 = Agak berbeda; 2 = Sangat berbeda; 9 = Variabel tidak dapat diterapkan pada situasi ini. Hanya kasus yang diidentifikasi oleh Butterworth sebagai “*core cases*” yang digunakan dan dianalisis; jika dependen variabel bernilai 9 (tidak dapat dipergunakan) maka kasus tidak disertakan. Data Butterworth dianalisis menggunakan program Turbo Pascal. Setelah menganalisis data dengan mempergunakan program jaringan saraf, Schrodt membandingkannya dengan empat metode lain: mode, ID3, analisis diskriminan, dan logit. Tabel I.2 menunjukkan nilai rata-rata dari akurasi dan ER statistik atas 10 percobaan dalam *neural networks*, 5 percobaan dalam analisis diskriminan dan analisis logit, serta 200 percobaan dalam ID3. Gambar I.3 menunjukkan hasil dari validasi.

Tabel I.1. Perhitungan dari Berbagai Metode

Statistic	Dependent variable				
	Stop	Abate	Isolate	Restrain	Settle
Accuracy					
Mode	.653	.586	.743	.545	.714
ID3	.581	.495	.663	.484	.603
Interval NN training	.647	.703	.830	.711	.752
Interval NN validation	.600	.536	.694	.523	.640
Binary NN training	.738	.798	.875	.724	.762
Binary NN validation	.604	.504	.724	.561	.660
Discriminant training	.918	.760	.931	.767	.779
Discriminant validation	.469	.543	.547	.523	.549
Logit training	.873	.775	.917	.741	.850
Logit validation	.613	.588	.654	.551	.658
Entropy ratio					
Mode	.320	.342	.303	.314	.300
ID3	.407	.401	.397	.407	.370
Interval NN training	.374	.589	.527	.597	.464
Interval NN validation	.320	.403	.374	.373	.302
Discriminant training	.921	.766	.937	.766	.792
Discriminant validation	.364	.499	.489	.490	.487
Logit validation	.521	.498	.494	.466	.459

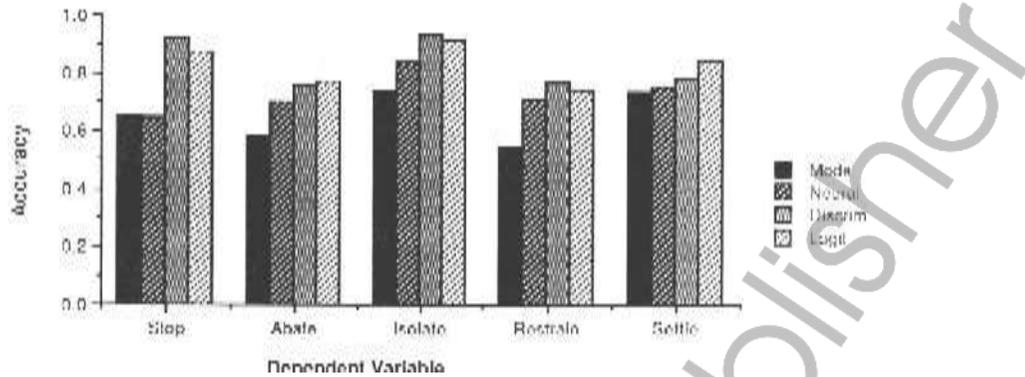
Sumber: Schrodt (1991)



Gambar I.3. Akurasi dalam *Validation Sets*
 Sumber: Schrodt (1991)

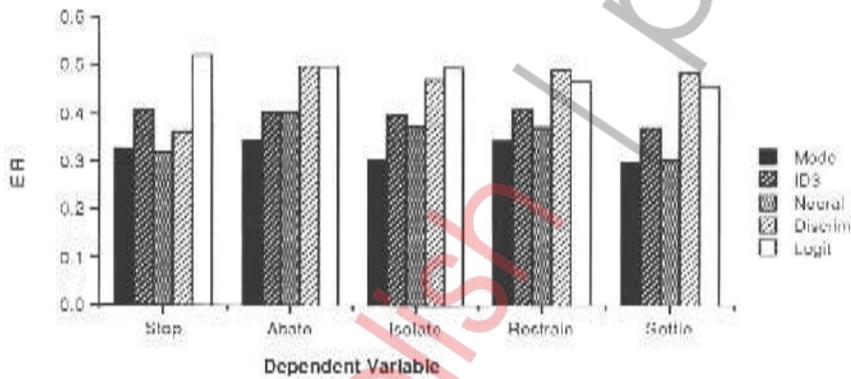
Dalam Gambar I.3. Jaringan saraf secara konsisten lebih akurat daripada model ID3; hal ini kira-kira sama dengan analisis diskriminan pada 2 dari 5 variabel dan secara substansial lebih akurat pada tiga variabel lainnya; dan memiliki akurasi yang hampir sama dengan model logit. Perbedaan antara akurasi jaringan saraf dan model diskriminan secara statistik signifikan pada tingkat 0,01 dalam uji-t untuk variabel Stop, Isolate, dan Settle. Namun demikian, keunggulan jaringan saraf hanya terlihat pada pengujian sampel-terpisah. Jika kasus pelatihan saja yang diuji, analisis diskriminan dan logit tampaknya memiliki keunggulan.

Gambar I.4 menunjukkan, keuntungan dari jaringan saraf hanya terlihat dalam tes sampel terpisah. Jika kasus pelatihan saja yang diuji, analisis diskriminan dan logit lebih unggul. Sementara diskriminan, cenderung kurang akurat pada set validasi. Sedangkan jaringan saraf, meski kurang akurat pada fit awal, lebih kuat daripada diskriminan sehubungan dengan kasus baru.



Gambar I.4. Akurasi dalam *Training Set*

Sumber: Schrodt (1991)



Gambar I.5. ER dalam *Validation Set*

Sumber: Schrodt (1991)

Analisis diskriminan dan logit jauh lebih baik daripada jaringan saraf atau ID3 dalam hal ukuran ini. Meskipun jaringan saraf sedikit lebih buruk daripada ID3, tetapi masih ada peningkatan sedikit dalam kemampuan prediksi. Namun, jaringan saraf memiliki hasil yang buruk dalam pengukuran ER, bahkan selama pelatihan. Skor ER jaringan saraf tidak pernah melebihi 0,60,

sementara analisis diskriminan dan logit mencapai 0,75 hingga 0,95. Jika jaringan saraf dilatih dalam waktu yang sangat lama, kemungkinan angka-angka tersebut dapat ditingkatkan.

Dari ekpreimentasi tersebut, Schrodt (1991) menemukan bahwa dalam memodelkan perilaku politik internasional, ternyata Jaringan saraf lebih akurat daripada model diskriminan, sedikit lebih akurat daripada model ID3, dan dapat dibandingkan dengan model logit ketika diuji dengan sampel data terpisah. Namun, ini tidak berarti jaringan saraf selalu lebih baik daripada model linier atau ID3. Data Butterworth memiliki karakteristik yang mempengaruhi kinerja jaringan saraf. Meskipun begitu, jaringan saraf tetap menjadi pesaing yang kuat.

Untuk saat ini, nilai utama jaringan saraf terletak pada kemampuannya untuk memprediksi perilaku dalam situasi dengan kebisingan dan data yang tidak lengkap. Dalam hal ini, mereka harus diperlakukan sebagai model linier umum dan digunakan dengan perhatian, pengabaian, atau kebijaksanaan yang sama, tergantung pada kasusnya. Dengan pengembangan lebih lanjut dalam konfigurasi jaringan dan algoritma pembelajaran, mungkin dimungkinkan untuk mencapai tingkat di mana model jaringan menjadi jelas lebih unggul daripada model linier, baik secara empiris maupun teoretis, dalam konteks metafora organisasi.

I.3.3. *Big Data* bagi Riset Prediktif

Guna menjelaskan bagaimana metode komputasi membantu dalam riset ekpserimentasi buku ini mempergunakan riset Gowa and Mansfield (1993) yang berjudul *Power Politics and International Trade* yang dimuat dalam *Jurnal American Political Science Association* tahun 1993. Gowa and Mansfield (1993) mempergunakan model gravitasi perdagangan bilateral yang biasanya dipergunakan untuk menjelaskan hubungan antara aliansi dan perdagangan bilateral. Riset dimulai dari keraguan Gowa

terhadap *prisoner dilemma* dalam mencerminkan aspek penting dari perjanjian perdagangan bebas dalam sistem internasional yang anarkis. Keraguan tersebut berangkat dari adanya *prisoner dilemma* atas keamanan eksternal. Bagi Gowa and Mansfield (1993), keamanan eksternal memiliki pengaruh penting dalam perdagangan bebas. Asumsi yang diajukan Gowa and Mansfield (1993) adalah (1) perdagangan bebas lebih mungkin terjadi di dalam aliansi politik-militer daripada antar aliansi; dan (2) aliansi cenderung menjadi koalisi perdagangan bebas jika terjadi dalam sistem bipolar daripada sistem multipolar.

Guna membuktikan asumsi tersebut, Gowa and Mansfield (1993) membangun dan menganalisis model permainan yang disebut dengan Model Gravitasi Perdagangan Bilateral (*Bilateral Gravity Model*). Dengan mempergunakan data dalam 80 tahun sejak tahun 1905, riset dimulai dengan menganalisis faktor ekonomi yang mempengaruhi perdagangan bilateral. Model melibatkan Produk Nasional Bruto (GNP), populasi negara importir dan eksportir, serta jarak geografis antara kedua negara. Berdasarkan penelitian sebelumnya yang menggunakan model ini, Gowa berharap bahwa nilai perdagangan bilateral secara langsung berkaitan dengan nominal GNP baik negara importir maupun eksportir, dan berbanding terbalik dengan jarak geografis di antara negara-negara tersebut. Selain GNP, populasi, dan jarak geografis, Gowa and Mansfield (1993) juga memasukkan dua variabel terkait aliansi dalam model, yaitu variabel yang menunjukkan apakah terdapat aliansi bilateral antara mitra dagang, dan variabel yang menunjukkan apakah terdapat aliansi multilateral di antara negara-negara tersebut. Variabel tambahan lain yang dimasukkan Gowa and Mansfield (1993) adalah variabel yang menunjukkan apakah importir dan eksportir sedang berperang atau tidak. Dengan demikian, model kami terdiri dari beberapa faktor yang

mempengaruhi perdagangan bilateral. Model yang disusun oleh Gowa and Mansfield (1993) adalah sebagai berikut:

$$\log X_{ij(t)} = \log A + B_1 \log Y_{i(t-1)} + B_2 \log Y_{j(t-1)} + B_3 \log P_{i(t-1)} + B_4 \log P_{j(t-1)} + B_5 \log D_{ij(t-1)} + B_6 \log BA_{ij(t-1)} + B_7 \log MA_{ij(t-1)} + B_8 \log War_{ij(t-1)} + \log Z_{ij}$$

Penjelasan:

- $\log X_{ij(t)}$ adalah nilai nominal ekspor (yang dinyatakan dalam dolar A.S.) oleh negara i ke negara j pada tahun t (pada tahun tersebut),
- $\log Y_{i(t-1)}$ adalah GNP nominal (yang dinyatakan dalam dolar A.S.) negara i pada tahun t - 1 (setahun sebelumnya sebelumnya),
- $\log Y_{j(t-1)}$ adalah GNP nominal (dinyatakan dalam dolar AS) negara j pada tahun t - 1 (setahun sebelumnya sebelumnya),
- $\log P_{i(t-1)}$ adalah populasi negara i pada tahun t - 1 (setahun sebelumnya sebelumnya),
- $\log P_{j(t-1)}$ adalah populasi negara j pada tahun t - 1 (setahun sebelumnya sebelumnya),
- $\log D_{ij(t-1)}$ adalah jarak geografis antara negara i dan j pada tahun t (pada tahun tersebut),
- $\log BA_{ij(t-1)}$ (variable dami) untuk menunjukkan apakah ada aliansi bilateral antara dua negara
- $\log MA_{ij(t-1)}$ untuk menunjukkan anggota perjanjian multilateral,
- $\log War_{ij(t-1)}$ untuk melihat apakah negara tersebut terlibat dalam perang dalam tahun sebelumnya,
- $\log z_{ij}$ adalah galat (error term).

Model yang dibangun Gowa and Mansfield (1993) tersebut menunjukkan bahwa nilai ekspor dipengaruhi oleh variabel-variabel tersebut. Hal ini juga berarti, dapat dipergunakan untuk mengamati apakah semua variabel di dalam model memiliki pengaruh dalam ekspor serta untuk melihat kekuatan pengaruh dari masing-masing variabel. Sementara itu, untuk memperoleh model yang baru, variabel-variabel tersebut diregresikan dengan hasil seperti yang ditunjukkan dalam Tabel I.1. dan Tabel I.2 di mana Tabel I.1. Gowa and Mansfield (1993) memasukkan Uni Soviet dan pada Tabel I.2. Gowa mengeluarkan Uni Soviet.

Tabel I.2. Regresi Ekspor pada GNP, Populasi, Jarak, Aliansi, dan Perang, 1905-1985

PARAMETER	PERIOD OF MULTIPOLARITY					PERIOD OF BIPOLARITY			
	1905	1913	1920	1930	1938	1955	1965	1975	1985
Intercept	-4.57 (7.88)	-8.79 (9.99)	57.21*** (14.74)	7.39 (5.06)	12.44* (6.19)	34.81*** (8.19)	5.69 (5.17)	6.29 (4.25)	12.14** (4.88)
log GNP _t	.95*** (.17)	1.68*** (.23)	2.78*** (.34)	1.53*** (.14)	1.67*** (.20)	1.12*** (.26)	.28 (.26)	.83*** (.20)	.96*** (.32)
log GNP _{t-1}	1.10*** (.18)	.90*** (.25)	2.17*** (.27)	1.25*** (.14)	1.57*** (.22)	.93*** (.25)	.44** (.26)	.55*** (.21)	1.19*** (.32)
log Population _t	-.02 (.33)	-.95*** (.35)	-4.10*** (.83)	-1.21*** (.23)	-1.68*** (.34)	-1.88*** (.49)	.14 (.39)	-.68*** (.26)	-1.13*** (.42)
log Population _{t-1}	-1.21*** (.27)	-.92*** (.31)	-4.83*** (.59)	-1.74*** (.23)	-2.27*** (.32)	-1.82*** (.49)	-.38 (.39)	-.45** (.25)	-1.42*** (.40)
log Distance _{ij}	-.33*** (.09)	-.06 (.10)	.27 (.13)	-.35*** (.06)	-.06 (.09)	-.01 (.09)	-.12** (.06)	-.23*** (.05)	-.28*** (.07)
log Bilat. alliance _{ij}	-.37 (.31)	-.57 (.36)	.20 (.77)	1.04*** (.42)	-.30 (.40)	3.02*** (.55)	2.58*** (.39)	2.07*** (.32)	2.10*** (.43)
log Multilat. alliance _{ij}	-.61 (.55)	-.31 (.72)	.96*** (.39)	— ^a	.48* (.35)	.86** (.46)	1.65*** (.33)	.99*** (.18)	.84*** (.26)
log War _{ij}	-7.12*** (.61)	— ^b	1.55 (.72)	— ^b	— ^b	— ^b	— ^b	— ^b	— ^b
Adjusted R ²	.92	.71	.83	.86	.80	.78	.82	.82	.80
N	39	39	37 ^c	40	37 ^d	41	40	41	41

Note: Entries are unstandardized regression coefficients with standard errors in parentheses. Years shown are year *t* in equation 1. For each year, there are 42 observations minus the number of outliers.

^aNo multilateral alliances existed among the major powers in 1929.

^bNo wars between major powers were conducted during these years.

^cNo data on the Soviet Union's exports to Germany are available for 1920.

^dNo data on Italian exports to the Soviet Union are available for 1938.

**p* ≤ .10 (one-tailed test); intercept *p* ≤ .10 (two-tailed test).

***p* ≤ .05 (one-tailed test); intercept *p* ≤ .05 (two-tailed test).

****p* ≤ .01 (one-tailed test); intercept *p* ≤ .01 (two-tailed test).

Sumber: Gowa and Mansfield (1993)

Data Uni Soviet dikeluarkan karena ada kemungkinan bahwa dengan memasukkan Uni Soviet dalam sampel akan melebih-lebihkan dampak aliansi pada perdagangan. Uni Soviet

memiliki komando dan ekonomi yang relatif mandiri selama ke-20. Kekuatan hubungan antara aliansi dan perdagangan mungkin dipengaruhi oleh sampel yang digunakan. Untuk memastikan hal tersebut, maka digunakan metode penaksiran kuadrat terkecil biasa untuk mendapatkan hasil yang tidak melibatkan Uni Soviet.

Tabel I.3. Regresi Ekspor pada GNP Per Kapita, Jarak, dan Aliansi (Kecuali Uni Soviet, 1905-1985)

PARAMETER	PERIOD OF MULTIPOLARITY				PERIOD OF BIPOLARITY			
	1905	1913	1920	1938	1955	1965	1975	1985
Intercept	9.57*** (2.54)	3.60 (3.46)	2.68 (2.95)	3.17 (3.05)	10.51*** (1.85)	8.73** (3.48)	8.33*** (2.79)	7.78* (4.27)
log Per capita GNP _t	1.12*** (.23)	1.78*** (.30)	1.88*** (.29)	1.41*** (.26)	.80*** (.21)	.79*** (.31)	.90*** (.21)	.69** (.30)
log Per capita GNP _{t-1}	.96*** (.23)	.69*** (.32)	.94*** (.28)	1.36*** (.26)	.51*** (.20)	.82*** (.30)	.78*** (.21)	1.15*** (.30)
log Distance _{it}	-.26** (.11)	.06 (.14)	-.21** (.10)	-.18** (.10)	-.14* (.10)	-.17* (.11)	-.21*** (.06)	-.32*** (.09)
log Bilat. alliance _{it}	-.41 (.39)	-.11 (.42)	.58 (.66)	-.21 (.62)	2.36*** (.49)	2.41*** (.51)	2.18*** (.32)	1.92*** (.44)
log Multilat. alliance _{it}	-.81 (.63)	-.12 (.75)	1.06** (.35)	.33 (.41)	1.31*** (.53)	.93* (.55)	.97*** (.25)	.80*** (.31)
Adjusted R ²	.75	.63	.74	.78	.87	.80	.85	.81
N	28	28	28	28	29	30	28	28

Note: Entries are unstandardized regression coefficients with standard errors in parentheses. Years shown are year *t* in equation 2. For each year, there are 30 observations minus the number of outliers.
 **p* ≤ .10 (one-tailed test); intercept *p* ≤ .10 (two-tailed test).
 ***p* ≤ .05 (one-tailed test); intercept *p* ≤ .05 (two-tailed test).
 ****p* ≤ .01 (one-tailed test); intercept *p* ≤ .01 (two-tailed test).

Sumber: Gowa and Mansfield (1993)

Gowa and Mansfield (1993) menyimpulkan bahwa kebijakan perdagangan bebas dapat menjadi kurang optimal bagi negara-negara dalam sistem internasional yang anarkis. Dalam beberapa kondisi, upaya untuk mendorong perdagangan dapat meningkatkan kesejahteraan. Hasil analisis empiris mendukung implikasi argumen Gowa tentang hubungan antara aliansi dan perdagangan. Hasil regresi menunjukkan bahwa aliansi menghasilkan efek langsung, signifikan secara statistik, dan arus perdagangan bilateral. Hasil regresi juga menunjukkan bahwa efek aliansi bervariasi. Aliansi bilateral terkadang memengaruhi arus perdagangan lebih kuat daripada aliansi multilateral; dan aliansi dalam sistem bipolar memiliki efek yang lebih kuat pada perdagangan daripada sistem multipolar.

I.4. Penutup

Big data memiliki peran yang sangat penting dan membantu dalam riset hubungan internasional. Dengan kemampuannya untuk mengumpulkan, menyimpan, dan menganalisis jumlah data yang besar dan kompleks, *big data* memberikan kesempatan baru untuk memahami dan menganalisis hubungan antara negara dan aktor internasional dengan cara yang lebih mendalam dan komprehensif.

Dorongan untuk mempelajari metode komputasi dalam riset hubungan internasional secara kuantitatif juga sangat penting. Dalam era digital ini, banyak data yang tersedia dan dapat diakses dengan cepat. Dengan mempelajari metode komputasi, seperti analisis data besar, statistik komputasi, dan teknik pengolahan bahasa alami, para peneliti hubungan internasional dapat memanfaatkan potensi besar yang ditawarkan oleh *big data* untuk mendapatkan wawasan yang lebih akurat dan mendalam tentang dinamika hubungan internasional. Metode komputasi juga membuka peluang baru untuk melakukan analisis kuantitatif yang lebih canggih, seperti pemodelan prediktif, analisis jaringan kompleks, dan analisis sentimen. Dengan menggabungkan kekuatan *big data* dengan metode komputasi ini, riset hubungan internasional dapat meningkat secara signifikan dalam hal akurasi, kecepatan, dan daya informasi.

Dengan demikian, penting bagi para peneliti dan akademisi yang tertarik dalam riset hubungan internasional untuk mempelajari metode komputasi dan memanfaatkan *big data*. Dengan memperkuat keterampilan dalam analisis kuantitatif dan pemanfaatan *big data*, mereka akan dapat menghasilkan penelitian yang lebih berkualitas, memberikan wawasan yang lebih mendalam, dan memberikan kontribusi yang lebih signifikan terhadap pemahaman kita tentang hubungan internasional.

Referensi

- Abbassi, Hanae, Imane El Alaoui, and Youssef Gahi. 2022. "Fraud Detection Techniques in the Big Data Era." In Scitepress, 161-70.
- Almeida, Fernando L F. 2017. "Benefits, Challenges and Tools of Big Data Management." *Journal of System Integration* 2017(4): 12-20.
- Al-Rawi, Ahmed, and Jacob Groshek. 2018. "Jihadist Propaganda on Social Media: An Examination of ISIS-Related Content on Twitter." *International Journal of Cyber Warfare and Terrorism* 8(4): 1-15.
- Balachandran, Bala M., and Shivika Prasad. 2017. "Challenges and Benefits of Deploying Big Data Analytics in the Cloud for Business Intelligence." In *Procedia Computer Science*, Elsevier B.V., 1112-22.
- Bhadani, A Kumar, and D Jothimani. 2016. "Big Data: Challenges, Opportunities and Realities." In *Effective Big Data Management and Opportunities for Implementation*, Pennsylvania, USA: IGI Global, 1-24.
- Bjola, Corneliu, and Lu Jiang. 2015. "Social Media and Public Diplomacy: A Comparative Analysis of the Digital Diplomatic Strategies of the EU, U.S. and Japan in China." In *Digital Diplomacy: Theory and Practice*, <https://www.researchgate.net/publication/286268365>.
- Chaney, Allison J. B., Brandon M. Stewart, and Barbara E. Engelhardt. 2018. "How Algorithmic Confounding in Recommendation Systems Increases Homogeneity and Decreases Utility." In *RecSys*, Association for Computing Machinery (ACM), 224-32.
- Cole, Dana, Jasmine Nelson, and Brian McDaniel. 2015. "Benefits and Risks of Big Data." In *Proceedings of the Southern Association for Information Systems Conference*, South Carolina:

- Association for Information Systems AIS Electronic Library (AISeL). <http://aisel.aisnet.org/sais2015/26>.
- Conte, R. *et al.* 2012. "Manifesto of Computational Social Science." *European Physical Journal: Special Topics* 214(1): 325–46.
- Edelmann, Achim, Tom Wolff, Danielle Montagne, and Christopher A Bail. 2020. *Computational Social Science and Sociology*.
- Edrington, Candice L., and Nicole M. Lee. 2018. "Tweeting a Social Movement: Black Lives Matter and Its Use of Twitter to Share Information, Build Community, and Promote Action." *Journal of Public Interest Communications* 2(2).
- Gandomi, Amir, and Murtaza Haider. 2015. "Beyond the Hype: Big Data Concepts, Methods, and Analytics." *International Journal of Information Management* 35(2): 137–44.
- Gowa, Joanne, and Edward D. Mansfield. 1993. "Power Politics and International Trade." *American Political Science Review* 87(2): 408–20.
- Hadi, Syed Ali. 2020. "Computational International Relations: A Dip in Digital Research." *Centre for Strategic and Contemporary Research*. <https://cscr.pk/explore/themes/social-issues/computational-international-relations-a-dip-in-digital-research/> (June 28, 2023).
- Haughton, Dominique, Mary Ann Robbert, Linda P Senne, and Vismay Gada. 2003. "Effect of Dirty Data on Analysis Results." In *Proceedings of the Eighth International Conference on Information Quality (ICIQ-03)*, <https://www.researchgate.net/publication/220918580>.
- IBM. "Big Data Analytic." <https://www.ibm.com/analytics/big-data-analyticsN.A>.
- Ijaz Baig, Maria, Liyana Shuib, and Elaheh Yadegaridehkordi. 2019. "Big Data Tools: Advantages and Disadvantages." *Journal of Soft Computing and Decision Support Systems* 6(6). <http://www.jscdss.com>.

- Kent, Michael L, and Maureen Taylor. 1998. "Building Dialogic Relationships Through the World Wide Web." *Public Relations Review* 24(3): 321-34.
- Kruikemeier, Sanne. 2014. "How Political Candidates Use Twitter and the Impact on Votes." *Computers in Human Behavior* 34: 131-39.
- Lamont, Christopher. 2015. *Research Methods in International Relations*. London: SAGE Publication Ltd.
- Melanie, Mitchell. 1999. *An Introduction to Genetic Algorithms*. 5th ed. London: A Bradford Book The MIT Press.
- Mojtahed, Vahid. 2019. "Big Data for Fraud Detection." In *Information and Communication Technologies (ICT) in Economic Modelling, Computational Social Sciences*, eds. F Cecconi and M Campenni. London: Avanade Ltd., 177-92.
- Molina, Mario, and Filiz Garip. 2019. *Machine Learning for Sociology*. New York. <https://doi.org/10.1146/annurev-soc-073117->
- Nelson, Laura K. 2020. "Computational Grounded Theory: A Methodological Framework." *Sociological Methods and Research* 49(1): 3-42.
- Phalapong, Pakorn. 2022. "Milk Tea Alliance: Constructing Transnational Identity by Consuming Milk Tea in the Asian Context 1 The Construction of Transnational Identity through the Consumption of Milk Tea in the Asian Context'." *Thai Journal of East Asian Studies* 26(2).
- Rahmadi. 2011. *Pengantar Metodologi Penelitian*. Banjarmasin: Antasari Press.
- Ridzuan, Fakhitah, and Wan Mohd Nazmee Wan Zainon. 2019. "A Review on Data Cleansing Methods for Big Data." In *Procedia Computer Science*, Elsevier B.V., 731-38.
- Salganik, Matthew J. 2018. *BIT BY BIT*. New Jersey: Princeton University Press.

- Schrodt, Philip A. 1991. "Prediction of Interstate Conflict Outcomes Using a Neural Network." *Social Science Computer Review* 9(3).
- Sevin, Efe, and Diana Ingenhoff. 2018. "Public Diplomacy on Social Media: Analyzing Networks and Content." *International Journal of Communication* 12.
- Shukla, Dadhichi, and Stephan Unger. 2022. "Sentiment Analysis of International Relations with Artificial Intelligence." *Athens Journal of Sciences* 9: 1-16. <https://doi.org/10.30958/ajs.X-Y-Z>.
- Miller, Taninecz Daniel, David Bachman, Steven Pfaff, and Jessica Beyer. 2019. "Three International Studies Computational Social Science Inquiries Examining Large Corpora of Natural Data." University of Washington.
- TechAmerica Foundation's Federal Big Data Commission. 2012. "Demystifying Big Data: A Practical Guide to Transforming the Business of Government." <http://www.techamerica.org/Docs/fileManager.cfm?f=techamerica-bigdatareport-final.pdf>. (June 29, 2023).
- Temby, Quinton. 2021. "The Milk Tea Alliance and China's Power Gap in Southeast Asia." *Perspective* 42(2021).
- Ünver, H. Akm. 2018. "Computational International Relations What Can Programming, Coding and Internet Research Do for the Discipline?" *All Azimuth: A Journal of Foreign Policy and Peace*. <https://dergipark.org.tr/en/doi/10.20991/allazimuth.476433>.
- Utama, Muhammad Anugrah. 2022. "Solidarity for Myanmar: #MilkTeaAlliance Indonesia's Activism in Pressuring the Government of Indonesia and ASEAN." *Global: Jurnal Politik Internasional* 24(2).

- Wolfsfeld, Gadi, Elad Segev, and Tamir Sheafer. 2013. "Social Media and the Arab Spring: Politics Comes First." *International Journal of Press/Politics* 18(2): 115-37.
- Zafar, Sadia *et al.* 2021. "Big Data: Challenges, Popular Tools Of Big Data-Benefits And Applications." *International Journal of Scientific & Technology Research* 10(5): 5. www.ijstr.org.

deepublish | publisher



Bab II

Text Mining dalam Penelitian Hubungan Internasional

Sebagaimana yang sudah kita pelajari pada Bab sebelumnya, metode komputasi menjadi salah satu alternatif bagi ilmuwan sosial dalam menjawab perkembangan jaman di mana *big data* memiliki peran yang sangat dominan dalam kehidupan manusia, pun demikian dalam ranah akademis. Data yang sudah tersimpan dalam internet atau media digital lainnya mungkin jumlahnya tidak terbatas, tetapi terkadang karena keterbatasan manusia, jumlah data yang tidak terbatas dan besar tersebut menjadi tidak bisa dimanfaatkan dengan baik. Oleh karena itu dalam bab ini menawarkan pengetahuan tentang salah satu jenis metode komputasional yang sangat menarik dan sering digunakan dalam riset-riset sosial, yaitu *text mining*. Secara singkat, *text mining* merupakan proses penambangan data dengan dibantu oleh perangkat lunak komputer untuk mendapatkan temuan yang diinginkan oleh ilmuwan.

Seperti dijelaskan pada bab sebelumnya, *posting-an* Twitter oleh publik figur, berita *online* dan berita di YouTube, teks pidato, hasil wawancara yang diunggah di internet merupakan beberapa data dari sekian banyak data yang bisa digunakan oleh ilmuwan sosial dalam riset-riset mereka. Tapi apakah anda membayangkan bagaimana jika data-data yang tidak terstruktur tersebut diolah secara manual? Pasti membutuhkan proses yang cukup memakan

waktu. Dengan menggunakan *text mining*, maka komputer yang akan membantu proses penambangan data-data tersebut. Tidak hanya menambang data, tetapi dalam *text mining* data yang tidak terstruktur akan dijadikan data yang lebih terstruktur untuk kemudian diolah lebih lanjut untuk memperoleh hasil yang diinginkan oleh ilmuwan.

Lalu, apa yang bisa menjadi temuan dalam *text mining*? Contoh sederhananya, ketika seorang ilmuwan HI ingin menganalisis sentimen publik terhadap wacana kepulauan mantan anggota Islamic State, maka melalui *text mining* temuan yang bisa di dapatkan adalah sentimen positif atau negatif yang dominan dalam persepsi publik terkait isu tersebut. Contoh lain adalah ketika seorang ilmuwan HI ingin menganalisis teks-teks pidato dari Barack Obama, maka melalui *text mining* akan didapatkan temuan berupa topik atau tema apa yang seringkali muncul dari teks-teks pidato tersebut. Selain itu, ilmuwan HI juga bisa mendapatkan temuan kata-kata apa yang sering muncul bersamaan dalam teks pidato tersebut. Itu adalah beberapa contoh temuan yang bisa didapatkan dengan menganalisis data menggunakan *text mining*.

Dengan semakin kompleksnya fenomena HI, dengan ketersediaan data yang semakin beragam dan jumlah yang sangat besar, tentu penggunaan metode *text mining* menjadi salah satu cara yang bisa digunakan oleh ilmuwan HI untuk menganalisis fenomena sosial yang sedang diperdebatkan. Pertama, bab ini akan menjelaskan tentang pengantar metode *text mining* yang memuat tentang definisi dan beberapa tahapan yang bisa dilakukan dalam *text mining*. Kedua, akan dijelaskan mengenai *Topic Modeling*—merupakan bagian dalam *text mining* yang digunakan untuk menganalisis teks untuk mendapatkan tema atau topik apa yang dominan. Ketiga, untuk lebih mudah memahami *text mining*, maka

akan diberikan contoh penerapan metode *text mining* dalam riset Hubungan Internasional.

II.1. Definisi Text Mining

Secara sederhana, *text mining* adalah proses untuk mendapatkan temuan dari teks (Kwartler, 2017). *Text mining* merupakan bagian dari *data mining*. Jika dilihat dalam skala yang lebih luas, *data mining* merupakan ekstraksi informasi atau pola menarik data dalam *database* yang besar dan terstruktur. Sementara itu, *text mining* sendiri merupakan prosedur menyintesis informasi, dengan menganalisis hubungan, pola, dan aturan di antara data tekstual menggunakan teknik dari *data mining*, *machine learning*, *natural language processing* (NLP), *information retrieval* (IR), dan *knowledge management* (Feldman & Sanger, 2007).

Jika mengacu pada pendapat para ahli, *text mining* bisa didefinisikan sebagai teknik menambang data berupa teks di mana sumber data berasal dari dokumen dan kemudian dilakukan analisis keterhubungan antar dokumen (Witten, 2004). Feldman dan Sanger mendefinisikan *text mining* sebagai suatu proses intensif pengetahuan di mana pengguna atau *user* berinteraksi dengan kumpulan dokumen dari waktu ke waktu dengan menggunakan seperangkat alat analisis (Feldman & Sanger, 2007). Dalam pandangan Feldman dan Sanger, proses penambangan data-penambangan teks-merupakan proses mengekstraksi informasi yang berguna dari sumber data melalui identifikasi dan eksplorasi pola yang menarik. Namun, dalam kasus penambangan teks, sumber datanya adalah kumpulan dokumen, dan pola menarik tidak ditemukan di antara catatan basis data yang diformalkan, tetapi di data tekstual yang tidak terstruktur dalam dokumen dalam kumpulan ini.

Dengan demikian, pada dasarnya *text mining* bertujuan untuk mengubah data teks yang tidak terstruktur menjadi data teks

yang lebih terstruktur. Data teks terstruktur tersebut nantinya bisa digunakan dalam tugas-tugas menambang data selanjutnya seperti pengklasifikasian dan pengelompokan (*clustering*). Sederhananya, *text mining* merupakan proses penambangan data berupa *text* yang tidak terstruktur dalam internet. Proses penambangan data tersebut dibantu oleh *software* yang berfungsi untuk mengidentifikasi pola, konsep, tema, kata kunci, dan lainnya yang kemudian berhubungan dengan *text* dalam data.

II.2. Kumpulan Dokumen dalam Metode *Text Mining*

Peran *big data* di era kontemporer ini menjadi semakin signifikan dalam berbagai sektor kehidupan manusia. Dari hari ke hari, ukuran data yang ada dalam *platform* digital menunjukkan peningkatan dengan kecepatan eksponensial (Talib *et al.*, 2016). Hal itu bisa ditunjukkan dengan hampir semua jenis institusi pemerintah, institusi pendidikan, industri bisnis, organisasi menyimpan datanya secara elektronik. Misalnya dalam institusi pendidikan, sebagian besar datanya sudah tersimpan dalam perpustakaan digital, repositori, media sosial, website, dan *platform* digital lainnya (Sagayam *et al.*, 2012). Perubahan tersebut tentu membawa pengaruh yang sangat signifikan pula pada pola akademisi dalam mengumpulkan dan mengolah data. Para akademisi juga dituntut oleh perkembangan jaman untuk menyesuaikan bagaimana mereka mampu mengolah data dalam jumlah yang besar. Hal itu tentunya menjadi tantangan karena cara menambang data secara tradisional tentu tidak mampu menangani jumlah data teks yang tak terbatas.

Pada pembahasan ini, fokus pembahasan metode komputasional yaitu data. Sebelum membahas mengenai *text mining* lebih jauh, kita akan fokus pada data. Menurut anda, apakah data berupa hasil wawancara—yang ada dalam *recorder* atau catatan pewawancara, *posting-an* presiden atau pejabat pemerintah di

media sosial, berita yang disiarkan di media elektronik atau *platform* digital seperti YouTube merupakan data terstruktur yang bisa langsung diolah dan diinterpretasikan? Mungkin anda akan menjawab ya jika data tersebut jumlahnya terbatas sehingga masih bisa diolah secara manual. Tetapi bayangkan jika data yang ada jumlahnya sangat besar, tentu menjadi lebih sulit untuk diolah dan diinterpretasikan secara manual. Dengan ketersediaan data yang semakin besar, maka kita membutuhkan alat atau perangkat untuk membantu mengolahnya sehingga kita sudah mendapatkan data jadi yang selanjutnya diinterpretasikan secara manual.

Dalam *text mining*, *document collection* (kumpulan dokumen) menjadi elemen yang sangat penting. Bisa dikatakan bahwa kumpulan dokumen merupakan seperangkat dokumen berbasis teks apapun. Tujuan *text mining* sendiri adalah untuk menemukan pola apa yang terbentuk dari kumpulan dokumen tersebut. Kumpulan dokumen ini jumlahnya sangat besar dapat berkisar ribuan bahkan puluhan juta. Kumpulan dokumen tersebut juga bersifat statis dan dinamis. Pada dokumen statis maka dokumen tidak akan berubah dari waktu ke waktu. Kebalikannya, pada dokumen yang dinamis maka dokumen tersebut dapat berubah, bertambah, dan disisipi oleh dokumen-dokumen baru lainnya dari waktu ke waktu (Feldman & Sanger, 2007). Perbedaan antara data statis dan dinamis tentu akan sangat berpengaruh pada pengolahan data dalam *text mining*. Semakin banyak dan semakin tinggi tingkat perubahan dokumen, maka akan menimbulkan tantangan pengoptimalan kinerja bagi berbagai komponen pada proses penambangan teks.

Selain itu perbedaan dari kumpulan dokumen juga bisa dibedakan menjadi data terstruktur, semi-terstruktur, dan tidak terstruktur. Data tidak terstruktur adalah data yang tidak memiliki format data yang teratur. Contoh data tidak terstruktur misalnya *tweet* dari para pengguna Twitter yang ditarik menggunakan

software (Misal python atau R studio). Data semi-terstruktur adalah data yang merupakan perpaduan antara format data terstruktur dan tidak terstruktur. Dalam hal ini, data semi terstruktur memiliki struktur yang cukup untuk memenuhi persyaratan basis data relasional tetapi data masih belum diolah lebih lanjut. Contoh data semi-terstruktur adalah data hasil survei, data hasil wawancara yang berupa video atau audio. Biasanya, data-data tersebut sudah memiliki struktur tetapi belum siap untuk dicari temuannya sehingga membutuhkan olah data lebih lanjut. Data terstruktur adalah data yang dibakukan ke dalam format tabel dengan banyak baris dan kolom, membuatnya lebih mudah untuk disimpan dan diproses sebagai dasar dari analisis data lebih lanjut. Contoh dari data terstruktur adalah data set dari Badan Pusat Statistik. Data tersebut dianggap terstruktur karena sudah jelas variabelnya, misal satu kolom sudah satu variabel, dan satu baris sudah satu observasi.

Dari perbedaan-perbedaan tersebut diharapkan dapat menambah pengetahuan dasar yang perlu dipahami sebelum masuk ke dalam pembahasan *text mining* lebih lanjut. Selain memahami perbedaan pada data yang biasa digunakan dalam *text mining*, ada fitur-fitur dokumen yang perlu dipahami. Fitur dokumen yang digunakan dalam penambangan teks memiliki dimensi yang jarang, artinya beberapa fitur muncul hanya dalam beberapa dokumen dan itu mempengaruhi pada pola yang terbentuk. Algoritma *text mining* juga bekerja pada representasi dokumen berbasis fitur, menciptakan kompromi antara penggambaran makna dokumen yang akurat dan efisiensi komputasi. Fitur-fitur dokumen yang umum digunakan adalah *characters*, *words*, *terms*, dan *concept* (Feldman & Sanger, 2007). Untuk memahami fitur-fitur tersebut maka akan dijelaskan di bawah ini:

- *Characters*: huruf individu, angka, dan karakter khusus yang membentuk fitur ke tingkat yang lebih tinggi seperti kata, istilah, dan konsep. Representasi berbasis karakter dapat mencakup semua karakter atau *subset*, dengan informasi posisi seperti bigram atau trigra.
- *Words*: Unit semantik dasar, dipilih langsung dari dokumen. Representasi dapat mencakup semua kata atau *subset* yang dioptimalkan, dengan menyaring kata penghubung dan karakter yang tidak bermakna.
- *Terms*: Kata tunggal dan frasa multi-kata diekstraksi dari dokumen menggunakan metode ekstraksi istilah. Representasi berbasis istilah terdiri dari istilah yang dipilih dari dokumen, menghasilkan representasi yang lebih kaya secara semantik daripada fitur tingkat kata.
- *Concept*: Fitur yang dihasilkan menggunakan metode kategorisasi manual, statistik, atau berbasis aturan. Fitur-fitur ini dapat menangkap kata tunggal, ekspresi, klausa, atau unit sintaksis yang lebih besar yang terkait dengan pengidentifikasi konsep tertentu.

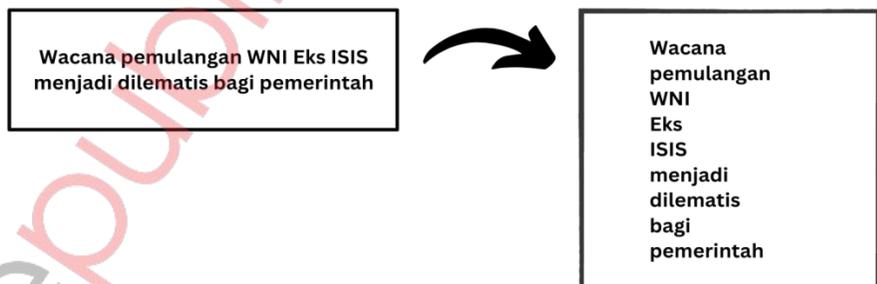
II.3. Pra-Pemrosesan dalam Metode *Text Mining*

Ada dua fase utama dalam metode *text mining*: 1) *Preprocessing* (pra-pemrosesan) dan integrasi data yang tidak terstruktur, dan 2) analisis statistik dari data yang diproses sebelumnya untuk mengekstraksi konten dari teks (Francis & Flynn, 2010). *Pra-processing* merupakan proses awal dalam menambang data untuk memastikan bahwa data yang telah didapatkan sudah siap untuk diolah lebih lanjut. Dalam tahapan ini, masih dibagi-bagi lagi menjadi beberapa sub tahapan di antaranya proses *tokenization*, *filtering*, *stemming*, dan *analyzing*. Untuk lebih jelasnya, tahapan dan sub tahapan tersebut dapat

dijelaskan lebih jauh. Dalam tahap *pra-processing*, sub tahapan yang bisa dilakukan bisa dijelaskan sebagai berikut:

a. *Tokenization*:

Tokenization merupakan langkah mendasar dalam proses penambangan teks atau kalimat yang kemudian dipecah menjadi unit individu yang disebut token. Token tersebut berupa kata, frasa, atau bahkan karakter, bergantung pada tingkat perincian yang diinginkan. Tokenisasi sendiri berfungsi sebagai dasar untuk banyak proses penambangan teks selanjutnya, seperti klasifikasi dokumen, analisis sentimen, dan pencarian informasi. Dengan memecah teks menjadi unit-unit yang lebih kecil ini memungkinkan penerapan teknik statistik dan komputasi untuk mengekstraksi makna, pola, dan wawasan dari data. Proses tokenisasi biasanya didasarkan pada pertimbangan tertentu dan disesuaikan dengan data yang dibutuhkan. Beberapa pertimbangan tersebut misalnya, pertama, tokenisasi berdasarkan tingkat kata yang menjadi bentuk paling umum dalam tokenisasi itu sendiri yaitu memecah teks menjadi kata-kata individu. Di bawah ini merupakan contoh dari tokenisasi yang didasarkan pada pecahan kata.



Gambar II.1. Proses Tokenisasi

Dari gambar di atas bisa dilihat bahwa kolom bagian kiri merupakan teks yang diinput untuk kemudian diproses untuk memecah kata dalam proses tokenisasi dan dalam kolom kanan terlihat hasil token dari kata-kata yang sudah berdiri secara individu.

Pertimbangan tokenisasi yang kedua adalah berdasarkan tanda baca atau karakter khusus. Dengan demikian, pertimbangan pecahan kalimat bisa dipengaruhi juga oleh tanda baca atau karakter khusus sebagai token terpisah atau menghapus semuanya, tergantung pada tujuan penelitiannya. Ketiga, tokenisasi bisa didasarkan pada penggunaan huruf kapital dan huruf kecil. Dalam beberapa kasus, tokenisasi melibatkan perubahan semua token menjadi huruf kecil untuk memastikan konsistensi dan menghindari memperlakukan kata yang sama secara berbeda karena kapitalisasi. Misalnya, "Apple" dan "apel" akan dianggap sebagai token yang sama. Dengan demikian penggunaan huruf kapital dan huruf kecil bisa sangat mempengaruhi data yang didapatkan, tetapi kembali lagi pada tujuan penelitiannya.

b. *Filtering*

Filtering merupakan proses menghilangkan informasi yang tidak diinginkan atau tidak relevan dari data teks. Jadi *filtering* bertujuan untuk menyaring data agar mendapatkan konten yang lebih relevan. Ada berbagai jenis teknik penyaringan yang digunakan dalam penambahan teks:

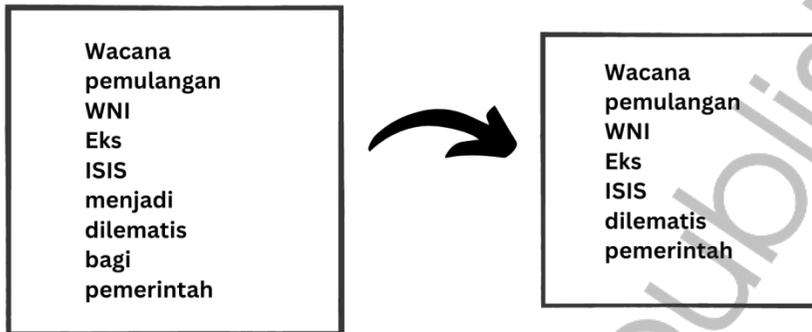
- Penghapusan *Stopword*: *Stopword* merupakan kata-kata umum dalam sebuah kalimat yang biasanya tidak memiliki kontribusi dalam analisis data. Kata-kata yang tidak memiliki arti yang signifikan tersebut contohnya adalah "dan," "itu," "adalah," "di", "ke", "dari", dan lain-lain. Kata-kata ini sering dihapus dari teks selama pemfilteran karena

dapat menambah kebisingan dan tidak berkontribusi banyak untuk analisis.

- Penghapusan Kebisingan: Data teks sering kali berisi elemen bising seperti *tag* HTML, karakter khusus, angka, atau tanda baca. Pemfilteran ini merupakan bertujuan untuk menghapus atau mengganti elemen-elemen ini untuk membersihkan teks sehingga meningkatkan kualitas data dan berimplikasi pada hasil yang lebih akurat.
- Pemfilteran khusus bahasa: merupakan teknik pemfilteran khusus yang melibatkan penanganan kontraksi, *stemming* atau *lemmatization* (mengurangi kata-kata ke bentuk dasarnya), atau berurusan dengan simbol atau pola khusus bahasa.
- Pengenalan Entitas: Pengenalan entitas melibatkan pengidentifikasian dan ekstraksi entitas tertentu atau entitas bernama dari teks, seperti nama orang, organisasi, lokasi, tanggal, dan lainnya. pemfilteran dapat dilakukan untuk fokus hanya pada entitas yang diekstraksi dan membuang bagian lain dari teks.
- Penyaringan Teks Panjang: Dalam beberapa kasus, dokumen teks dapat difilter berdasarkan panjangnya. Dokumen pendek atau sangat panjang dapat dikecualikan dari analisis untuk memastikan kumpulan data yang seimbang dan representatif.
- Penyaringan Khusus Domain: Bergantung pada domain atau konteks tertentu dari proses penambangan teks. Teknik pemfilteran khusus domain dapat diterapkan dalam penghapusan istilah, frasa, atau pola tertentu yang tidak relevan dengan analisis.

Dari teknik penyaringan data yang telah dijelaskan di atas, jelas bahwa proses pemfilteran merupakan proses penting yang harus dilakukan sebelum masuk pada olah data lebih lanjut untuk

bisa menyaring data menjadi data yang lebih relevan sehingga mendapatkan hasil yang lebih akurat. Di bawah ini merupakan contoh dari proses filterisasi:



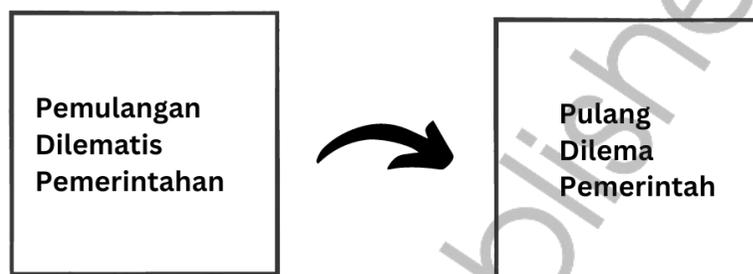
Gambar II.2. Proses Filterisasi

Dari contoh di atas, dapat dilihat bahwa proses *filtering* telah dilakukan dengan menghapus kata-kata yang dianggap tidak relevan untuk olah data lebih lanjut seperti “menjadi” dan “bagi”. Dengan demikian, kata-kata yang disisakan hanya kata-kata yang dianggap berguna untuk olah data lebih lanjut.

c. *Stemming*

Dalam penambangan teks, *stemming* adalah proses mereduksi kata menjadi bentuk dasar atau akarnya, yang dikenal sebagai "*stem*". Akar kata mewakili makna inti dari sebuah kata dan diturunkan dengan menghilangkan awalan, akhiran, dan variasi lain yang menunjukkan bentuk tegang, jamak, atau gramatikal. Tujuan dari *stemming* adalah untuk mengonsolidasikan bentuk kata yang berbeda menjadi bentuk dasar yang sama, sehingga variasi dari kata yang sama diperlakukan sebagai tanda yang identik selama analisis. Ini membantu mengurangi dimensi data dan

meningkatkan akurasi tugas penambangan teks tertentu. Berikut adalah contoh untuk mengilustrasikan *stemming*:



Gambar II.3. Proses *Stemming*

Dari ilustrasi di atas dapat diidentifikasi bahwa kata-kata yang memiliki imbuhan di-*stem* dan menjadikan kata tersebut diturunkan kembali ke akarnya dengan menghilangkan imbuhan "an" dan "tis" dari kata aslinya. *Stemming* dapat bermanfaat dalam tugas penambangan teks seperti pengambilan informasi, pengelompokan dokumen, dan pemodelan topik, karena membantu mengurangi ketersebaran data dan mengonsolidasikan kata-kata yang mirip. Namun, penting untuk diperhatikan bahwa *stemming* adalah teknik penyederhanaan dan dapat menyebabkan hilangnya presisi, karena bentuk kata tertentu mungkin memiliki arti yang berbeda dalam konteks yang berbeda.

II.4. *Topic Modeling*

Setelah melalui tahapan *pre-processing*, tahapan selanjutnya yang harus dilakukan dalam *text mining* adalah *analyzing*. *Analyzing* merupakan proses menganalisis data untuk mendapatkan *insight* atau temuan. Salah satu cara yang bisa digunakan pada proses menganalisis data adalah *topic modeling*. *Topic modeling* merupakan teknik yang mengungkapkan, menemukan, dan menandai struktur tematik dalam kumpulan dokumen dengan menggunakan

algoritma tertentu di dalam *software* (Kherwa & Bansal, 2020). Sederhananya, *topic modeling* merupakan metode untuk menemukan pola dalam suatu kumpulan *text* berdasarkan kemunculan kata dan pola keberulangan untuk mendapatkan tema dari *text* tersebut. Dengan demikian, pada *topic modeling*, *text* tersebut dikelompokkan menjadi beberapa topik berdasarkan tema yang sudah didapatkan. *Topic modeling* menjadi penting untuk dilakukan karena dalam metode ini dapat menemukan pola penggunaan kata dan cara menghubungkan dokumen yang memiliki pola serupa. Dokumen yang dimaksud adalah campuran topik, di mana topik adalah distribusi probabilitas kata-kata. Dengan demikian dalam *topic modeling* bisa didapatkan probabilistik sederhana di mana dokumen dapat dihasilkan (Alghamdi & Alfalqi, 2015)

Dalam pendekatan *topic modeling*, ada beberapa metode yang bisa dilakukan di antaranya *Latent Semantic Analysis* (LSA), *Latent Dirichlet Allocation* (LDA), *Probabilistic Latent Semantic Analysis* (pLSA), dan lain-lain (Nurlyayli & Nasichuddin, 2019). LDA menjadi salah satu metode yang sangat populer digunakan dalam *topic modeling*. Demikian pula dengan riset-riset sosial yang lebih sering menggunakan metode LDA. Dengan demikian, pada pembahasan ini akan fokus pada metode LDA. LDA adalah model probabilitas yang digunakan untuk menemukan topik tersembunyi dalam sebuah koleksi dokumen. Model ini mengasumsikan bahwa setiap dokumen adalah campuran dari berbagai topik, dan setiap topik adalah distribusi kata-kata. Ide dasarnya adalah bahwa dokumen-dokumen dihasilkan dengan menggabungkan topik-topik yang berbeda, dan setiap kata dalam dokumen dipilih berdasarkan topik terkaitnya. Secara sederhana, LDA bertujuan untuk mengungkap struktur tersembunyi topik dalam korpus dokumen. Dengan menganalisis pola dan kemunculan kata-kata, LDA membantu kita mengidentifikasi topik-topik yang mendasar yang ada dalam

dataset (Silge & Robinson, 2017). Berikut ini adalah pemahaman yang lebih rinci tentang konsep-konsep utama dalam LDA:

- **Dokumen:** Korpus terdiri dari kumpulan dokumen. Dokumen-dokumen ini bisa berupa artikel, pos blog, makalah penelitian, ulasan pelanggan, pos media sosial, atau bentuk data teks lainnya.
- **Topik:** Setiap dokumen diasumsikan sebagai campuran dari berbagai topik. Topik dapat dianggap sebagai tema atau subjek yang muncul dalam sekelompok dokumen. Misalnya, dalam kumpulan artikel berita, topik-topik dapat berupa "politik," "olahraga," "hiburan," dan sebagainya.
- **Kata-kata:** Kata-kata adalah unit dasar teks dalam dokumen. LDA mengasumsikan bahwa setiap kata dalam dokumen terkait dengan salah satu topik. Pemilihan kata dalam sebuah dokumen dipengaruhi oleh campuran topik yang ada dalam dokumen tersebut.
- **Distribusi Topik:** LDA mengasumsikan bahwa terdapat distribusi topik secara keseluruhan dalam korpus dokumen. Artinya, beberapa topik mungkin lebih dominan daripada yang lain dalam koleksi dokumen tersebut.
- **Distribusi Kata:** LDA juga mengasumsikan bahwa setiap topik memiliki distribusi kata-kata yang terkait dengannya. Distribusi kata-kata ini menentukan probabilitas kemunculan kata tertentu dalam sebuah dokumen berdasarkan topiknya.

Dengan asumsi-asumsi ini, LDA bertujuan untuk mengungkapkan distribusi topik untuk setiap dokumen dan distribusi kata-kata untuk setiap topik. Dengan menganalisis distribusi ini, kita dapat memperoleh wawasan tentang topik-topik yang dominan dalam korpus dan bagaimana kata-kata terkait dengan topik-topik tersebut. Secara ringkas, Langkah 1 dalam pemodelan topik LDA melibatkan pemahaman konsep dasar LDA,

termasuk asumsi-asumsi tentang hubungan antara dokumen, topik, dan kata-kata. Pemahaman ini menjadi dasar untuk langkah-langkah berikutnya, di mana kita akan menjelajahi cara mempersiapkan data, membangun model LDA, dan menginterpretasi hasil untuk menghasilkan topik-topik yang bermakna dari kumpulan dokumen (Silge & Robinson, 2017).

II.5. Penerapan *Text Mining* menggunakan *Topic Modeling* dalam Riset Hubungan Internasional

Setelah memahami dasar konseptual dari *topic modeling*, maka dalam sub bab ini akan dijelaskan contoh penerapannya dalam riset Hubungan Internasional untuk memudahkan pembaca dalam memahami penggunaan *topic modeling*. Mengutip dari penelitian oleh (Schoenfeld *et al.*, 2018), penelitian tersebut bertujuan untuk menganalisis pidato yang dibuat oleh negara-negara anggota Dewan Keamanan PBB tentang Afganistan. Tujuan lainnya adalah untuk memahami bagaimana negara-negara-khususnya negara anggota DK PBB-berbicara tentang situasi di Afganistan dan topik apa yang paling penting bagi mereka. Dalam penelitiannya, para penulis menggunakan metode campuran (*mix-method*) untuk mempelajari lanskap diskursif dari perdebatan di DK PBB tentang Afganistan. Dengan metode tersebut, korpus besar pidato yang dibuat oleh DK PBB (termasuk anggota tetap, anggota rotasi, tamu, dan pejabat PBB) dari awal intervensi di Afganistan hingga tahun 2017 dikumpulkan. Tahapan selanjutnya, digunakan *Latent Dirichlet Allocation (LDA)* untuk mengidentifikasi topik apa yang dibahas dalam pidato dan untuk menetapkan setiap pidato ke satu atau lebih topik. Selain LDA, juga digunakan *speaker-topic network* untuk memvisualisasikan hubungan antara orator dan topik-*network analysis* akan dijelaskan lebih lanjut pada Bab IV. Dengan demikian, contoh dalam pembahasan ini hanya difokuskan pada penggunaan metode LDA

Setelah melakukan olah data dengan metode LDA, para penulis mengidentifikasi beberapa topik utama yang dibahas dalam pidato tersebut, di antaranya adalah topik-topik seperti keamanan, pemerintahan, pembangunan, dan kerja sama regional. Yang menarik, hasil temuan juga memperlihatkan bahwa relevansi topik tersebut berubah dari waktu ke waktu, dengan keamanan menjadi topik yang paling menonjol di tahun-tahun awal intervensi, disusul oleh topik tentang tata kelola dan pembangunan yang menjadi lebih penting di tahun-tahun berikutnya. Tidak hanya terbatas pada topik apa yang menonjol, tetapi para penulis juga mengidentifikasi beberapa aktor kunci, termasuk Amerika Serikat, Afganistan, dan Pakistan, dan menemukan bahwa hubungan mereka dengan aktor lain berubah seiring waktu. Secara keseluruhan, hasil penelitian ini menunjukkan bahwa menggunakan program komputer untuk menganalisis pidato dapat memberikan wawasan baru ke dalam politik internasional dan dapat membantu peneliti untuk lebih memahami bagaimana negara berbicara tentang isu-isu penting. Di bawah ini merupakan contoh topik-topik hasil temuan dengan menggunakan metode LDA.

Tabel II.1. Temuan Topik dengan Metode LDA

	Topic	Characterizing key-words⁴	Peak year(s)
T1	Elections	elections, process, electoral, security, election, political, government, Japan, presidential, international.	Main attention 2004-2005; some attention 2009-2010, 2012-2014.
T2	Development	Afghanistan, development, international, efforts, security, community, secretary, assistance, general, afghan.	Main attention 2006-2009.
T3	Ceremonial 1 (Undefined)	will, council, one, time, need, must, many, members, us, even.	Little attention overall.
T4	Drugs	drug, Afghanistan, drugs, trafficking, narcotics, production, counter, opium, cultivation, crime.	Clear peaks 2001 and 2003.
T5	National Peace Process	peace, government, support, process, afghan, Afghanistan, unity, national, conflict, attacks.	Main attention 2015-2017.
T6	Transition	Afghanistan, afghan, will, transition, process, European, union, conference, regional, support.	Main attention 2010-2014.
T7	Women/ Human Rights	Afghanistan, women, rights, security, government, afghan, human, national, progress, 2014.	Main attention 2011-2016.
T8	Reconstruction	reconstruction, police, administration, ISAF, Bonn, Jirga, million, Afghanistan, assistance, security.	Clear peak in 2002, some attention until 2006.
T9	Regional Security	Afghanistan, Pakistan, people, Taliban, peace, terrorism, Iran, India, security, international	Main attention 2001, 2015-2017; some attention overall.
T10	Ceremonial 2 (Broad Issues)	united, nations, afghan, will, must, support, new, work, also, international.	Main attention 2007-2011.

Tabel di atas memberikan ringkasan rinci dari sepuluh topik yang diidentifikasi oleh para penulis dengan menggunakan metode LDA. Tabel ini dibagi menjadi empat kolom. Kolom pertama mencantumkan sepuluh topik, yang diberi label T1 hingga T10.

Kolom kedua memberikan deskripsi singkat dari setiap topik, yang membantu pembaca memahami tema utama yang muncul dalam pidato. Misalnya, Topik 1 dijelaskan sebagai "Elections" sedangkan Topik 2 dijelaskan sebagai "Development", dan seterusnya. Artinya adalah dalam suatu pidato dalam rentang waktu tertentu, maka topik mengenai "Pemilu" dan "Pembangunan" menjadi lebih menonjol dibanding topik yang lain. Sementara itu dalam kolom ketiga lebih dipertegas mengenai kata-kata teratas apa yang muncul untuk mendefinisikan setiap topik dominan. Pada T1 bisa dilihat bahwa kata-kata yang muncul mendefinisikan topik mengenai "Elections" adalah "elections", "process", "electoral", "security", "election", "political", "government", "Japan", "presidential", "international". Kata-kata tersebut diidentifikasi oleh algoritma LDA sebagai kata-kata paling menonjol dalam setiap topik, dan memberikan ringkasan yang berguna dari tema utama yang muncul dalam pidato. Pada kolom keempat menyediakan data mengenai rentang waktu di mana suatu topik menjadi lebih ditonjolkan dalam suatu pidato. Data tersebut membantu pembaca dalam memahami bahwa relevansi dalam setiap topik dalam berubah dari waktu ke waktu. Bisa dilihat dalam tabel T2 bahwa isu mengenai "Pembangunan" menjadi menonjol dibicarakan dalam pidato pada rentang waktu antara tahun 2006-2009. Sementara relevansi topik mengalami perubahan yang ditunjukkan T6, di mana topik yang dominan menarik perhatian dalam pidato adalah topik tentang "Transisi".

Dari contoh di atas, dapat dikatakan bahwa penggunaan *topic modeling* dengan metode LDA bisa menjadi alternatif dalam melakukan analisis teks untuk menemukan topik-topik apa yang dominan dari waktu ke waktu terkait suatu isu yang sedang hangat diperdebatkan oleh publik baik dalam bentuk teks pidato atau bahkan teks-teks dalam media sosial seperti Twitter. Penggunaan metode ini jelas memudahkan para peneliti sosial khususnya ketika

mereka berhadapan dengan data dalam jumlah besar dan juga data yang tidak teratur.

II.6. Penutup

Penggunaan *text mining* dalam menganalisis fenomena Hubungan Internasional dewasa ini menjadi salah satu alternatif untuk memperkaya metode riset selain menggunakan perspektif-perspektif yang *mainstream*. Dengan bantuan *software* yang memiliki algoritma-algoritma tertentu, maka data yang sangat besar dan kompleks dapat diolah dengan lebih mudah untuk mendapatkan hasil temuan yang diinginkan oleh peneliti. Metode *text mining* sendiri memiliki aspek kelebihan dan kekurangan ketika digunakan dalam penelitian sosial, termasuk dalam penelitian Hubungan Internasional, di antaranya

1. Efisiensi: Metode *text mining* dapat memproses data dalam jumlah besar dengan cepat dan efisien, yang dapat menghemat waktu dan sumber daya peneliti.
2. Objektivitas: Metode *text mining* dapat membantu peneliti mengidentifikasi pola dan tren dalam data tanpa pengaruh bias pribadi atau interpretasi subjektif.
3. Fleksibilitas: Metode *text mining* dapat diterapkan pada berbagai sumber data, termasuk media sosial, artikel berita, dan dokumen pemerintah.
4. Wawasan: Metode *text mining* dapat mengungkapkan wawasan dan koneksi yang mungkin tidak langsung terlihat melalui metode penelitian tradisional.

Sementara itu, kekurangan dalam metode *text mining* adalah:

1. Akurasi: Metode *text mining* mungkin tidak selalu memberikan hasil yang akurat, terutama saat bekerja dengan data tidak terstruktur atau data yang sulit dikategorikan.

2. Bias: Metode *text mining* dapat dipengaruhi oleh bias dari algoritma atau model yang digunakan, yang dapat menyebabkan hasil yang kurang akurat
3. Privasi: Metode *text mining* dapat menimbulkan kekhawatiran tentang privasi dan perlindungan data, terutama saat bekerja dengan data sensitif atau pribadi.
4. Interpretasi: Metode *text mining* dapat menghasilkan hasil yang sulit untuk diinterpretasikan atau memerlukan analisis tambahan untuk dipahami sepenuhnya.

Secara keseluruhan, metode *text mining* dapat menjadi metode riset yang untuk mengkaji penelitian sosial berbasis *big data*. Namun, dalam implementasinya, peneliti perlu untuk lebih menyadari potensi keterbatasan dan tantangan terkait penggunaan metode ini. Dengan demikian, untuk menutup aspek kekurangan dalam olah data yang mungkin bisa terjadi dalam metode *text mining*, biasanya metode ini akan dilengkapi dengan penggunaan metode penelitian lain (*mix-methods*). Dengan demikian, temuan yang didapatkan dari olah data akan menjadi lebih akurat dan tidak bias.

Referensi

- Alghamdi, R., & Alfalqi, K. (2015). A Survey of Topic Modeling in Text Mining. *International Journal of Advanced Computer Science and Applications*, 6(1), 7. http://thesai.org/Downloads/Volume6No1/Paper_21-A_Survey_of_Topic_Modeling_in_Text_Mining.pdf
- Feldman, R., & Sanger, J. (2007). *The text mining handbook : advanced approaches in analyzing unstructured data*. Cambridge University Press.
- Francis, L., & Flynn, M. (2010). *Text Mining Handbook*.
- Kherwa, P., & Bansal, P. (2020). Topic Modeling: A Comprehensive Review. *EAI Endorsed Transactions on Scalable Information*

- Systems*, 7(24), 1–16. <https://doi.org/10.4108/eai.13-7-2018.159623>
- Kwartler, T. (2017). Text Mining in Practice with R. In *Text Mining in Practice with R*. <https://doi.org/10.1002/9781119282105>
- Nurlayli, A., & Nasichuddin, M. A. (2019). Topik Modeling Penelitian Dosen Jptei Uny Pada Google Scholar Menggunakan Latent Dirichlet Allocation. *Elinvo (Electronics, Informatics, and Vocational Education)*, 4(2), 154–161. <https://doi.org/10.21831/elinvo.v4i2.28254>
- Sagayam, R., Srinivasan, S., & Roshni, S. (2012). A Survey of Text Mining: Retrieval, Extraction and Indexing Techniques. *International Journal Of Computational Engineering Research (Ijceronline.Com)*, 2(5), 2250–3005. <http://pakacademicsearch.com/pdf-files/com/319/1443-1446>
Volume 2, Issue 5, September, 2012.pdf
- Schoenfeld, M., Eckhard, S., Patz, R., & van Meegdenburg, H. (2018). *Discursive Landscapes and Unsupervised Topic Modeling in IR: A Validation of Text-As-Data Approaches through a New Corpus of UN Security Council Speeches on Afghanistan*. October. <http://arxiv.org/abs/1810.05572>
- Silge, J., & Robinson, D. (2017). Welcome to Text Mining with R. In *Development*.
- Talib, R., Kashif, M., Ayesha, S., & Fatima, F. (2016). Text Mining: Techniques, Applications and Issues. *International Journal of Advanced Computer Science and Applications*, 7(11), 414–418. <https://doi.org/10.14569/ijacsa.2016.071153>
- Witten, I. H. (2004). Text mining. *The Practical Handbook of Internet Computing*, 14-1-14-22. <https://doi.org/10.1201/9780203507223>

Bab III

Analisis Sentimen

Analisis sentimen, juga dikenal sebagai penambangan opini, adalah teknik teks *mining* yang melibatkan penentuan sentimen emosional atau subjektif yang terungkap dalam sebuah teks. Tujuannya adalah memahami dan menguantifikasi kesan emosional dari sebuah dokumen, apakah itu positif, negatif, netral, atau bahkan emosi spesifik seperti kegembiraan, kemarahan, kesedihan, atau terkejut (Silge & Robinson, 2017). Setiap usaha yang melibatkan analisis emosi atau sentimen yang diungkapkan dalam teks dapat mendapatkan manfaat dari analisis sentimen. Analisis sentimen adalah alat yang sangat berguna yang dapat memberikan wawasan berharga tentang nada emosional, sikap, dan pendapat yang terungkap dalam data teks (Appiahene *et al.*, 2022).

Contoh penggunaan sentimen analisis yang paling umum adalah dalam analisis *review* atau umpan balik yang diberikan pelanggan. Analisis sentimen memiliki peran yang penting dalam menganalisis umpan balik dari pelanggan atau pengguna. Dalam konteks bisnis, analisis sentimen dapat memberikan wawasan yang berharga tentang sikap dan opini pelanggan terhadap produk, layanan, dan merek. Dengan menganalisis umpan balik pelanggan di media sosial, situs ulasan, dan *platform* lainnya, bisnis dapat mengidentifikasi area yang perlu diperbaiki dan membuat keputusan berdasarkan data. Misalnya, dengan mendeteksi sentimen negatif terhadap fitur produk tertentu, bisnis dapat mengambil langkah untuk memperbaikinya atau menghapusnya

sepenuhnya. Di sisi lain, dengan mendeteksi sentimen positif terhadap kampanye pemasaran tertentu, bisnis dapat mengadopsinya dalam kampanye masa depan (Devika *et al.*, 2016).

Selain itu, analisis sentimen juga membantu dalam manajemen pengalaman pelanggan. Dengan memantau umpan balik pelanggan di media sosial dan *platform* lainnya, bisnis dapat dengan cepat mendeteksi umpan balik negatif dan mengambil tindakan untuk menangani masalah tersebut. Dengan berinteraksi langsung dengan pelanggan yang tidak puas, bisnis dapat mengubah pengalaman pelanggan yang buruk menjadi pengalaman yang positif dan meningkatkan kepuasan pelanggan secara keseluruhan. Dalam hal ini, analisis sentimen memberikan wawasan yang berharga tentang preferensi dan kebutuhan pelanggan, yang dapat membantu bisnis dalam meningkatkan produk, layanan, dan proses bisnis mereka.

Dengan menggabungkan analisis sentimen dalam menganalisis umpan balik pelanggan, bisnis dapat memiliki pemahaman yang lebih mendalam tentang sentimen dan perasaan pelanggan mereka. Ini memungkinkan mereka untuk mengambil tindakan yang tepat dan proaktif dalam meningkatkan produk, layanan, dan pengalaman pelanggan secara keseluruhan. Kegunaan lain analisis sentimen adalah untuk mengukur opini publik berdasarkan sosial media. Analisis sentimen dapat digunakan oleh departemen pemerintah untuk mengukur sentimen publik terhadap layanan dan kebijakan mereka. Dengan menganalisis komentar di *platform* seperti Facebook dan Twitter, mereka dapat mengidentifikasi area yang perlu diperbaiki dan membuat keputusan yang berdasarkan informasi untuk meningkatkan layanan publik. Misalnya, jika sebuah departemen pemerintah mendeteksi sentimen negatif terhadap layanan publik tertentu, mereka dapat mengambil langkah untuk memperbaikinya atau

mengalokasikan lebih banyak sumber daya ke layanan tersebut (Devika *et al.*, 2016).

III.1. Metode dan Teknik Analisis Sentimen

Ada berbagai cara dan teknik yang dapat digunakan untuk melakukan analisis sentimen. Berbagai metode ini berangkat dari pandangan metodologis yang berbeda, dan beberapa memiliki fungsi dan tujuan yang berbeda juga.

III.1.1. Pendekatan Machine Learning

Pendekatan *machine learning* melibatkan penggunaan teknik pembelajaran mesin untuk mengklasifikasikan sentimen. *Machine learning* adalah cara bagi komputer untuk belajar dan membuat prediksi tanpa diprogram secara langsung. Ini seperti mengajar komputer untuk berpikir dan membuat keputusan sendiri. Daripada memberikan komputer daftar aturan yang harus diikuti, kita memberikannya banyak contoh dan membiarkannya menemukan pola dan hubungan secara mandiri. Bayangkan Anda memiliki program komputer yang perlu menganalisis ulasan pelanggan tentang sebuah produk untuk menentukan apakah ulasan tersebut positif atau negatif. Alih-alih membaca setiap ulasan secara manual dan mengklasifikasikannya, Anda dapat menggunakan *machine learning* untuk mengajari komputer melakukannya.

Pertama, Anda mengumpulkan *dataset* besar berisi ulasan pelanggan, di mana setiap ulasan diberi label sebagai positif, negatif, atau netral. Anda memberikan *dataset* ini ke dalam algoritma *machine learning* dan membiarkannya belajar dari pola dan hubungan dalam data. Algoritma menganalisis kata-kata, frasa, dan konteks dari setiap ulasan dan mencoba menemukan pola yang menunjukkan sentimen positif atau negatif. Misalnya, algoritma dapat belajar bahwa kata-kata seperti "bagus," "luar biasa," dan "suka" sering dikaitkan dengan sentimen positif, sementara kata-

kata seperti "buruk," "kecewa," dan "mengerikan" sering dikaitkan dengan sentimen negatif. Setelah model *machine learning* dilatih dengan data ini, Anda dapat memberikan ulasan pelanggan baru yang belum pernah dilihat sebelumnya. Model akan menggunakan pola yang dipelajari selama pelatihan untuk memprediksi sentimen dari setiap ulasan. Misalnya, jika Anda memberikan ulasan baru yang menyatakan, "Saya benar-benar menyukai produk ini!", model kemungkinan akan mengklasifikasikannya sebagai sentimen positif.

Dengan terus memperbaiki model dan memberikannya lebih banyak data yang diberi label, Anda dapat meningkatkan akurasi dari waktu ke waktu. Semakin banyak data dan contoh yang Anda berikan, semakin baik model tersebut dalam mengklasifikasikan sentimen dengan akurat pada teks baru. Dengan cara ini, *machine learning* memungkinkan Anda untuk mengotomatiskan analisis sentimen, membuatnya lebih cepat dan efisien dibandingkan analisis manual. Ini memungkinkan Anda untuk memproses volume teks yang besar dan mendapatkan wawasan berharga tentang pendapat dan umpan balik pelanggan. Pendekatan *machine learning* sering menghasilkan akurasi yang tinggi, terutama ketika menggunakan teknik seperti *deep learning* dan *neural networks*. Namun, penting untuk memahami bahwa model yang dilatih pada satu domain atau bahasa mungkin tidak bekerja dengan baik pada domain atau bahasa lain.

III.1.2. Pendekatan Rule Based

Analisis sentimen berbasis aturan adalah metode untuk menganalisis sentimen dalam teks dengan menggunakan aturan yang telah ditentukan sebelumnya. Dalam pendekatan ini, kita membuat seperangkat aturan atau pedoman yang digunakan untuk mengklasifikasikan sentimen dalam teks. Pendekatan *rule based* menggunakan aturan-aturan yang telah ditentukan sebelumnya untuk mengklasifikasikan sentimen. Aturan ini bisa berupa aturan

linguistik yang melibatkan analisis tata bahasa, sintaksis, dan struktur kalimat, atau aturan berbasis logika yang menggunakan pernyataan-pernyataan yang spesifik untuk mengklasifikasikan sentimen. Contoh penggunaan analisis sentimen berbasis aturan adalah dengan menggunakan aturan linguistik. Misalkan kita ingin menganalisis sentimen dalam kalimat "Saya sangat senang dengan layanan pelanggan yang diberikan oleh perusahaan ini." Dalam aturan linguistik, kita dapat menentukan bahwa kata-kata seperti "senang" dan "layanan pelanggan" mengindikasikan sentimen positif. Oleh karena itu, berdasarkan aturan ini, kita dapat mengklasifikasikan kalimat tersebut sebagai sentimen positif.

Selain aturan linguistik, analisis sentimen berbasis aturan juga dapat menggunakan aturan berbasis logika. Misalkan kita memiliki aturan berbasis logika yang menyatakan bahwa jika suatu kalimat mengandung kata-kata negatif seperti "tidak" atau "bukan", maka sentimen dalam kalimat tersebut dianggap negatif. Sebagai contoh, jika kita memiliki kalimat "Produk ini tidak memenuhi harapan saya", aturan tersebut akan mengklasifikasikan kalimat tersebut sebagai sentimen negatif. Pendekatan analisis sentimen berbasis aturan dapat digunakan untuk mengklasifikasikan sentimen dalam teks dengan cepat dan mudah. Namun, metode ini memerlukan pembuatan aturan secara manual dan perlu diperbarui secara berkala sesuai dengan perubahan dalam bahasa atau tren baru.

Misalnya, sebuah aturan bisa mengatakan bahwa jika suatu kalimat mengandung kata-kata positif seperti "baik" atau "senang", maka sentimennya adalah positif. Pendekatan ini melibatkan pemisahan setiap kalimat dalam dokumen menjadi kata-kata individu (tokenisasi) dan menguji setiap token atau kata untuk mencocokkan aturan-aturan yang ada. Pendekatan *rule based* dapat efektif dalam mengklasifikasikan sentimen, tetapi memerlukan

pembuatan dan pemeliharaan aturan secara manual, yang bisa memakan waktu dan sumber daya manusia.

III.1.3. Analisis Semantik

Pendekatan analisis semantik mempertimbangkan makna dan konteks teks dalam analisis sentimen. Daripada menganalisis kata demi kata, pendekatan ini menggunakan evaluasi n-gram. N-gram adalah pola kata yang terdiri dari beberapa kata yang berdekatan dalam teks. Misalnya, bigram adalah pola dua kata yang berdekatan, sedangkan trigram adalah pola tiga kata yang berdekatan. Dengan menggunakan n-gram, pendekatan ini mencoba memahami makna keseluruhan dari pola kata tersebut. Analisis semantik mempertimbangkan konteks yang lebih luas dan berusaha memasukkan semantik dalam proses analisis sentimen. Pendekatan ini dapat menggunakan teknik seperti *latent semantic analysis* dan *topic modeling* untuk mengidentifikasi pola dan tren dalam data tanpa adanya label sentimen.

Analisis sentimen semantik adalah metode untuk menganalisis sentimen dalam teks dengan mempertimbangkan makna dan konteks kata-kata. Pendekatan ini melibatkan pemahaman lebih dalam tentang teks, bukan hanya melihat kata per kata. Misalkan kita memiliki kalimat "Film ini sangat menarik dan menghibur." Dalam analisis sentimen semantik, kita tidak hanya melihat kata "menarik" dan "menghibur" secara terpisah, tetapi juga mempertimbangkan hubungan antara kedua kata tersebut. Dalam hal ini, kita dapat menyimpulkan bahwa kalimat ini memiliki sentimen positif, karena kata-kata "menarik" dan "menghibur" memiliki konotasi positif.

Selain itu, analisis sentimen semantik juga dapat mempertimbangkan konteks yang lebih luas dalam teks. Misalkan kita memiliki kalimat "Saya tidak senang dengan pelayanan buruk yang saya terima di restoran ini." Dalam analisis sentimen semantik, kita dapat memperhatikan kata-kata "tidak senang" dan

"pelayanan buruk" sebagai indikasi sentimen negatif, meskipun kata "tidak" sebenarnya memiliki konotasi negatif. Dengan mempertimbangkan konteks dan makna keseluruhan kalimat, analisis sentimen semantik dapat menghasilkan hasil yang lebih akurat.

III.1.4. Pendekatan Berbasis Leksikon

Analisis sentimen berbasis leksikon adalah salah satu pendekatan dalam memahami sentimen dalam teks dengan menggunakan leksikon atau daftar kata sentimen yang telah ditentukan sebelumnya. Leksikon sentimen terdiri dari kata-kata yang diberi skor sentimen, biasanya dalam bentuk skala dari positif hingga negatif.

Pada dasarnya, pendekatan ini melibatkan langkah-langkah berikut:

1. **Pembentukan Leksikon Sentimen:** Leksikon sentimen dikembangkan melalui proses penelitian yang melibatkan penilaian manusia terhadap kata-kata. Para peneliti mengevaluasi apakah sebuah kata memiliki makna positif, negatif, atau netral. Kata-kata ini kemudian dikumpulkan dalam leksikon sentimen yang dapat digunakan untuk analisis sentimen. Dalam banyak kasus, analisis sentimen dapat dilakukan dengan menggunakan leksikon yang telah dikembangkan oleh orang lain. Ini yang membuat analisis sentimen berbasis leksikon merupakan salah satu metode analisis sentimen yang paling sederhana dan mudah digunakan.
2. **Penilaian Kata-kata dalam Teks:** Dalam analisis sentimen berbasis leksikon, setiap kata dalam teks dianalisis secara individual. Kata-kata tersebut dicocokkan dengan entri dalam leksikon sentimen untuk mengetahui skor sentimen mereka. Jika sebuah kata ditemukan dalam leksikon, skor

sentimen yang sesuai diberikan. Jika kata tidak ditemukan, dapat dianggap netral atau metode lain seperti pendekatan jarak terdekat dapat digunakan untuk menentukan skor sentimen yang sesuai.

3. Agregasi Skor Sentimen: Setelah semua kata dalam teks dinilai, skor sentimen untuk setiap kata dijumlahkan atau diagregasikan untuk menentukan sentimen keseluruhan dari teks. Ini dapat dilakukan dengan menghitung total skor, mengambil rata-rata skor, atau menggunakan metode lain sesuai dengan kebutuhan dan tujuan analisis.

Contoh penggunaan analisis sentimen berbasis leksikon adalah sebagai berikut:

Misalkan kita memiliki leksikon sentimen yang memasukkan kata "bagus" dengan skor +1 dan kata "buruk" dengan skor -1. Ketika menganalisis kalimat "Film ini sangat bagus, tetapi aktingnya buruk", kita menemukan kata "bagus" yang memiliki skor +1 dan kata "buruk" yang memiliki skor -1 dalam leksikon. Dalam hal ini, dengan mengagregasikan skor sentimen, kita dapat menyimpulkan bahwa sentimen keseluruhan kalimat tersebut adalah netral ($+1 - 1 = 0$). Meskipun ada kata-kata dengan sentimen yang berlawanan, agregasi skor sentimen memberikan pemahaman yang lebih akurat tentang sentimen keseluruhan kalimat.

Pendekatan analisis sentimen berbasis leksikon memiliki beberapa kelebihan. Pertama, pendekatan ini relatif mudah diimplementasikan karena hanya memerlukan leksikon atau daftar kata sentimen yang telah ada. Kedua, karena skor sentimen telah ditentukan sebelumnya, pendekatan ini dapat memberikan hasil yang konsisten. Namun, pendekatan ini juga memiliki beberapa kelemahan. Pertama, leksikon sentimen perlu diperbarui secara rutin untuk mempertahankan relevansi dengan perubahan bahasa dan tren baru. Kedua, pendekatan ini mungkin tidak dapat

mengatasi kata-kata atau frasa yang memiliki makna ganda atau konteks yang kompleks. Misalnya, jika sebuah kata memiliki makna positif dalam satu konteks tetapi makna negatif dalam konteks lain, analisis sentimen berbasis leksikon mungkin tidak dapat mengatasi ambiguitas ini.

Analisis sentimen berbasis leksikon adalah pendekatan yang menggunakan leksikon atau daftar kata sentimen yang telah ditentukan sebelumnya untuk menganalisis sentimen dalam teks. Dengan memberikan skor sentimen pada kata-kata dalam teks dan mengagregasikannya, analisis sentimen berbasis leksikon dapat memberikan pemahaman tentang sentimen keseluruhan dalam teks. Penting untuk memperhatikan kelebihan dan kelemahan pendekatan ini serta mempertimbangkan konteks dan tujuan analisis untuk memperoleh hasil yang lebih akurat.

III.2. Contoh Analisis Sentimen secara Praktis (Menggunakan Leksikon)

Untuk melakukan analisis sentimen pada teks pendek seperti ulasan dan *tweet*, Anda dapat mengikuti proses langkah demi langkah. Pertama, kumpulkan kumpulan data ulasan pelanggan, misalnya dari sebuah restoran. Ulasan-ulasan ini mungkin mencakup pernyataan seperti "Makanannya enak sekali!" atau "Pelayanannya lambat dan mengecewakan." Selanjutnya, lakukan pra-pemrosesan data dengan membersihkannya, mengubah teks menjadi huruf kecil, mengubah kata menjadi kata dasar dan/atau kata baku, serta menghapus unsur yang tidak perlu seperti tanda baca.

Langkah berikutnya adalah tokenisasi. Tokenisasi adalah proses membagi teks menjadi unit-unit terkecil yang disebut "token". Dalam contoh "Makanannya enak sekali", proses tokenisasi akan memisahkan kalimat tersebut menjadi token-token yang terdiri dari kata-kata individu:

1. "Makanannya"
2. "enak"
3. "sekali"

Dalam tokenisasi, setiap kata dalam kalimat diberikan token sendiri. Ini membantu kita untuk menganalisis teks dengan lebih terperinci, memahami makna kata secara terpisah, atau melakukan penghitungan frekuensi kata. Tokenisasi dapat dilakukan menggunakan beberapa metode atau algoritma yang sesuai dengan bahasa yang digunakan dan tujuan analisis. Metode umum yang digunakan termasuk pemisahan berdasarkan spasi, pemisahan berdasarkan tanda baca, atau penggunaan algoritma N-gram yang lebih kompleks (misal, "enak sekali" dianggap satu token.)

Salah satu langkah penting berikutnya dalam pra-pemrosesan teks adalah menghapus *stopwords*. *Stopwords* adalah kata-kata umum dalam bahasa yang tidak banyak memberikan kontribusi pada sentimen keseluruhan atau makna sebuah kalimat. Contoh *stopwords* dalam bahasa Indonesia termasuk "dan," "di," "dari," "itu," dan sebagainya. Kata-kata ini sering muncul tetapi seringkali tidak membawa banyak sentimen atau konteks.

Setelah menghapus *stopwords*, Anda dapat membuat leksikon sentimen atau kamus yang terdiri dari kata-kata atau frasa beserta skor sentimen yang terkait. Namun, perlu dicatat bahwa membuat leksikon tidak selalu diperlukan. Ada leksikon yang sudah ada yang telah disusun oleh peneliti atau ahli analisis sentimen lainnya. Leksikon ini berisi skor sentimen yang telah ditentukan sebelumnya untuk kata-kata dan dapat digunakan langsung dalam analisis Anda. Dengan memanfaatkan leksikon yang sudah ada, Anda dapat menghemat waktu dan upaya dalam mengembangkan leksikon sendiri. Selanjutnya, berikan skor sentimen pada token-token yang tersisa dalam teks dengan mencari mereka dalam leksikon. Misalnya, kata "enak" mungkin memiliki skor sentimen positif +1.

Untuk mengilustrasikan hal ini, mari berikan skor sentimen pada token-token dalam ulasan "Makanannya enak sekali!". Kita dapat menggunakan leksikon sentimen sederhana dengan tiga kategori sentimen: positif (+1), negatif (-1), dan netral (0).

Tabel III.1. Contoh Penetapan Skor Sentimen untuk "Makanannya Enak Sekali!"

Token	Skor Sentimen
Makanan	0
Enak	+1
Sekali	0

Dalam contoh ini, skor sentimen untuk setiap token dalam ulasan ditentukan berdasarkan leksikon sentimen. Kata "Makanan" bersifat netral dan diberi skor sentimen 0. Namun, kata "enak" dikaitkan dengan skor sentimen positif +1. **S**

Untuk menentukan sentimen keseluruhan, agregasikan skor sentimen dari semua token dalam teks. Hal ini dapat dilakukan dengan menjumlahkan skor atau mengambil rata-ratanya. Terapkan nilai ambang batas untuk mengklasifikasikan sentimen sebagai positif, negatif, atau netral berdasarkan skor yang telah diagregasi. Misalnya, jika skor di atas 0, beri label sebagai sentimen positif. Jika skor di bawah 0, beri label sebagai sentimen negatif. Jika tidak, beri label sebagai sentimen netral. Misal, dalam kalimat di atas skor total adalah +1, maka kalimat itu dapat dikategorikan sebagai kalimat dengan sentimen positif. Berikut adalah contoh lain analisis sentimen untuk kalimat "Pelayanannya buruk tapi makanannya enak sekali":

Tabel III.2. Contoh Penetapan Skor Sentimen untuk "Pelayanannya Buruk tapi Makanannya Enak Sekali"

Kata	Skor Sentimen
Pelayanannya	0
Buruk	-1
Tapi	0
Makanannya	0
Enak	+1
Sekali	0

Dalam contoh di atas, kata "pelayanannya", "makanannya", dan "sekali" diberi skor sentimen 0 karena tidak memiliki sentimen yang kuat. Kata "buruk" diberi skor sentimen -1 karena menunjukkan sentimen negatif terkait dengan pelayanan. Kata "enak" diberi skor sentimen +1 karena menggambarkan pengalaman yang positif terkait dengan makanan. Total nilai sentimen untuk kalimat ini kemudian menjadi 0 (-1 + 1), sehingga kalimat ini dapat dikategorikan dalam sentimen netral.

III.3. Contoh Analisis Sentimen menggunakan Pembelajaran Mesin (*Supervised Learning*)

Kita dapat menggunakan pendekatan pembelajaran terawasi untuk melatih komputer secara otomatis menganalisis dan memprediksi sentimen dari teks pendek. Mari kita ikuti langkah-langkah yang terlibat:

Pertama, kita perlu mengumpulkan *dataset* teks pendek dan sentimen yang sesuai. Misalnya, kita dapat mengumpulkan ulasan pelanggan untuk sebuah restoran dan memberi label sentimen positif, negatif, atau netral. Selanjutnya, kita mempersiapkan data dengan membersihkan teks. Ini melibatkan penghapusan simbol dan tanda baca, mengonversi teks menjadi huruf kecil, dan memisahkan kata-kata. Sebagai contoh, frasa "Makanannya enak sekali!" menjadi "makanannya enak sekali." Untuk lebih memahami kata-kata dalam teks, kita mengidentifikasi *stopwords*. Ini adalah

kata-kata umum tapi tidak memiliki nilai atau makna sentimental, seperti "dan," "yang," atau "tetapi."

Supervised learning adalah metode mengajari komputer untuk mempelajari pola tertentu. Untuk belajar, komputer memerlukan "bahan" yang bisa dipelajari. Bahan ini perlu kita siapkan. Salah satu cara untuk mempersiapkan bahan ini adalah melalui pelabelan manual.

Tabel III.3. Pelabelan Manual

Teks	Sentimen
Makanannya enak sekali!	Positif
Pelayanannya buruk tapi makanannya enak sekali	Netral
Pengalaman yang menyenangkan	Positif
Pelayanannya sangat lambat	Negatif
Restorannya kotor dan tidak terawat	Negatif

Dalam tabel ini, kita memberikan label sentimen positif, negatif, atau netral untuk setiap teks. Misalnya, teks "Makanannya enak sekali!" diberi label sentimen positif karena menyatakan kepuasan terhadap makanan. Teks "Pelayanannya buruk tapi makanannya enak sekali" diberi label sentimen netral karena memiliki sentimen positif dan negatif yang saling meniadakan. Begitu seterusnya.

Tabel ini kemudian kita berikan kepada komputer untuk dipelajari. Tanpa berusaha masuk ke dalam ranah yang terlalu teknis, pada dasarnya komputer akan mengenali pola mana-mana *tweet* yang positif dan mana yang negatif. Pengetahuan pola ini kemudian dapat digunakan untuk secara otomatis memberikan penilaian sentimen pada teks-teks yang lain. Kita menguji pembelajaran komputer dengan memberikannya teks baru dalam bahasa Indonesia, seperti "Pelayanannya buruk tapi makanannya enak sekali." Kita meminta komputer untuk memprediksi sentimennya. Jika komputer dengan benar memprediksi sentimen

positif, itu menunjukkan bahwa ia telah belajar dari contoh-contoh pelatihan. Untuk meningkatkan akurasi prediksi komputer, kita dapat menambahkan bahan pembelajaran untuk komputer.

Setelah komputer belajar dan meningkat, kita dapat menggunakan model yang telah dilatih untuk menganalisis sentimen dalam teks baru yang belum pernah dilihat sebelumnya dalam bahasa Indonesia, seperti ulasan pelanggan atau *posting* media sosial tentang sebuah restoran. Model tersebut akan memprediksi apakah sentimen tersebut positif, negatif, atau netral. Dengan mengikuti langkah-langkah ini, kita dapat memanfaatkan pembelajaran terawasi untuk melakukan analisis sentimen pada teks pendek, membantu kita memahami sentimen yang terungkap dalam teks berbahasa Indonesia.

III.4. Visualisasi Hasil Sentimen Analisis

Ada banyak cara untuk memvisualisasi hasil analisis sentimen. Secara umum, visualisasi hasil sentimen analisis dilakukan untuk menggambarkan kata-kata yang bernilai positif dan negatif. Salah satu cara yang paling umum digunakan adalah menggunakan *word cloud* atau awan kata. Awan kata merupakan kumpulan kata yang digambarkan dalam bentuk kumpulan seperti awan. Ukuran dari masing-masing kata menggambarkan jumlah kata dalam teks itu. Dalam konteks sentimen, biasanya dibuat dua *word cloud* berbeda untuk kategori kata positif dan kata negatif.

III.5. Penggunaan Analisis Sentimen dalam Riset Hubungan Internasional

Salah satu contoh penggunaan analisis sentimen dalam riset HI adalah dalam pengambilan keputusan. Georgiadou *et al.* (2020) menjelajahi kemungkinan penggunaan analisis sentimen dari data Twitter untuk membantu pengambilan keputusan pemerintah dalam negosiasi Brexit antara Inggris dan Uni Eropa. Para penulis berpendapat bahwa analisis sentimen dapat memberikan informasi waktu nyata tentang pendapat publik terhadap hasil negosiasi, sehingga membantu negosiator menghindari kegagalan kebijakan dan negosiasi.

Tujuan penelitian ini adalah menunjukkan manfaat analisis sentimen dalam meningkatkan pengambilan keputusan dan strategi dalam kebijakan publik dan negosiasi. Secara khusus, para penulis ingin menunjukkan bagaimana analisis sentimen dapat digunakan untuk mengidentifikasi hasil yang disukai dan yang tidak disukai dalam negosiasi internasional, serta memetakan preferensi masyarakat terkait hasil tersebut secara perbandingan.

Para penulis mengumpulkan dan menganalisis lebih dari 1,5 juta *tweet* terkait negosiasi Brexit dari Mei hingga November 2018. Mereka menggunakan teknik pemrosesan bahasa alami dan pembelajaran mesin untuk mengklasifikasikan sentimen setiap *tweet* menjadi positif, negatif, atau netral. Selanjutnya, mereka menggambarkan fluktuasi sentimen dari waktu ke waktu dan menghubungkannya dengan peristiwa penting dalam negosiasi. Hasil ini dapat digunakan untuk mengidentifikasi kapan negosiasi mendapatkan dukungan dari publik dan kapan negosiasi mendapatkan penentangan.

Penelitian ini menunjukkan potensi analisis sentimen dalam memberikan umpan balik waktu nyata tentang pendapat publik terhadap kebijakan tertentu. Analisis sentimen dapat mengidentifikasi kecenderungan preferensi masyarakat terhadap kebijakan dalam jangka waktu tertentu. Para penulis juga

menyebutkan bahwa analisis sentimen dapat membantu mengungkapkan hasil yang disukai dan yang tidak disukai dalam negosiasi internasional, baik pada waktu tertentu maupun selama periode waktu tertentu. Penelitian ini menekankan nilai analisis sentimen dalam pengambilan keputusan dan strategi dalam kebijakan publik dan negosiasi.

Dalam diskusi, para penulis membahas bagaimana analisis sentimen dapat membantu memahami hasil yang disukai dan yang tidak disukai dalam negosiasi internasional, membantu pembuat kebijakan dan negosiator memahami batasan skenario yang diinginkan. Mereka juga menyoroti bahwa analisis sentimen dapat menjadi alat yang berguna sebagai tambahan dari jajak pendapat tradisional, yang seringkali mahal dan memakan waktu. Analisis sentimen dapat memberikan informasi waktu nyata yang efisien tentang pendapat publik.

III.6. Penutup

Penggunaan analisis sentimen dalam ilmu sosial memang menawarkan berbagai potensi dan manfaat yang besar. Namun, seperti halnya dengan setiap alat analisis, ada juga tantangan dan permasalahan yang perlu dihadapi dalam penerapannya. Beberapa tantangan dan permasalahan penting yang terkait dengan penggunaan analisis sentimen dalam ilmu sosial antara lain:

1. Ambiguitas dan kompleksitas teks: Analisis sentimen sering mengandalkan pemrosesan teks yang dapat menjadi rumit. Teks seringkali memiliki struktur yang kompleks, dan arti *sentiment* yang sebenarnya dapat bergantung pada konteks, konotasi, dan *figurative language* yang digunakan. Memahami dan menginterpretasi sentimen dengan tepat bisa menjadi sulit.
2. Tidak ada standar emosi: Tidak ada kesepakatan universal tentang definisi emosi tertentu atau tentang bagaimana mengukur emosi secara akurat. Setiap individu dapat

mengalami dan mengekspresikan emosi dengan cara yang berbeda-beda. Hal ini membuat sulit untuk mengembangkan algoritma atau model yang dapat mengenali emosi dengan tingkat keakuratan yang tinggi.

3. Subjektivitas dan keberagaman: Pengalaman emosi dan interpretasi *sentiment* dapat sangat subjektif. Setiap individu memiliki latar belakang, budaya, dan konteks unik yang mempengaruhi persepsinya terhadap sentimen tertentu. Oleh karena itu, analisis sentimen sering kali melibatkan banyak keberagaman dalam data dan interpretasi.
4. Data yang tidak terstruktur: Data sosial yang digunakan untuk analisis sentimen seringkali tidak terstruktur dan beragam dalam format. Teks yang ada dalam media sosial, forum *online*, atau catatan kualitatif sering kali tidak teratur, tidak terstruktur, atau tidak gramatikal. Memproses data semacam ini menjadi tantangan tersendiri dalam analisis sentimen.
5. Pengaruh dan bias dalam analisis: Meskipun analisis sentimen didesain untuk objektivitas, mereka masih rentan terhadap pengaruh dan bias. Proses *preprocessing*, pengembangan model, dan interpretasi hasil analisis dapat mencerminkan bias dari data pelatihan atau keputusan yang dibuat oleh pembuatnya. Kualitas analisis sentimen juga dapat dipengaruhi oleh cara pemilihan data dan pemberian label sentimen.

Dalam menghadapi tantangan ini, penting untuk memahami keterbatasan dan konteks penggunaan analisis sentimen dalam ilmu sosial. Penerapannya harus diimbangi dengan pemahaman mendalam tentang subjektivitas, kompleksitas teks, dan perbedaan individu. Dengan pemahaman yang matang, analisis sentimen dapat memberikan wawasan berharga dalam

memahami emosi dan sentimen manusia dalam konteks sosial yang lebih luas.

Referensi

- Appiahene, P., Afrifa, S., Akwa Kyei, E., & Nimbe, P. (2022). Understanding the Uses, Approaches and Applications of Sentiment Analysis. University of Energy and Natural Resources.
- Devika, M. D., Sunitha, C., *et al.* (2016). Sentiment Analysis: A Comparative Study on Different Approaches. *Procedia Computer Science*, 87, 1.
- Georgiadou, E., Angelopoulos, S., *et al.* (2020). Big data analytics and international negotiations: Sentiment analysis of Brexit negotiating outcomes. *International Journal of Information Management*, 51, 4.
- Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, 5(1), 1-167.
- Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends® in Information Retrieval*, 2(1-2), 1-135.
- Silge, J., & Robinson, D. (2017). *Text mining with R: A tidy approach*. O'Reilly Media.



Bab IV

Analisis Jaringan dalam Ilmu HI

Analisis jaringan telah menjadi pendekatan baru dalam penelitian ilmu politik, terutama dalam bidang hubungan internasional (HI). Bedanya dengan cara tradisional mengatur organisasi seperti pasar atau hierarki negara, analisis jaringan memberikan pandangan yang lebih luas dan berbeda. Jaringan merupakan kumpulan hubungan yang membentuk struktur, yang pada gilirannya dapat membatasi dan memungkinkan para pelaku. Dengan analisis jaringan, kita memiliki teori dan alat yang dapat digunakan untuk mencari tahu pola-pola struktur jaringan dan menguji hipotesis-hipotesis tentang struktur ini. Di sisi lain, pasar dan hierarki negara dijelaskan sebagai hubungan tawar-menawar yang sementara dan birokrasi formal, yang tidak dapat menangkap kompleksitas struktur jaringan.

Pentingnya analisis jaringan dalam hubungan internasional terbukti dalam kemampuannya untuk menggambarkan dengan tepat jaringan internasional, menyelidiki dampak jaringan terhadap hasil-hasil internasional kunci, dan kemampuan aktor untuk meningkatkan kekuasaan mereka dengan meningkatkan dan memanfaatkan posisi mereka dalam jaringan. Analisis jaringan menawarkan seperangkat alat untuk mengidentifikasi dan mengukur sifat struktural jaringan serta sekumpulan teori yang menghubungkan struktur dengan hasil-hasil. Pendekatan ini menantang pandangan konvensional tentang kekuasaan dalam

hubungan internasional dengan mendefinisikan kekuasaan jaringan dalam tiga cara yang berbeda: akses, perantara, dan opsi keluar.

Peneliti HI seharusnya menggunakan analisis jaringan karena ia menyediakan seperangkat teori dan alat untuk menciptakan tantangan dan menguji proposisi tentang struktur jaringan. Analisis jaringan memungkinkan penyelidikan dan pengukuran struktur jaringan, yang merupakan sifat-sifat yang muncul dari pola hubungan yang berkelanjutan antara para pelaku yang dapat menentukan, memungkinkan, dan membatasi perilaku-pelaku tersebut. Pendekatan ini juga menantang pandangan konvensional tentang kekuasaan dalam hubungan internasional dengan mendefinisikan kekuasaan jaringan dalam tiga cara yang berbeda: akses, perantara, dan opsi keluar. Nilai analisis jaringan dalam hubungan internasional terbukti dalam deskripsi yang tepat tentang jaringan internasional, penyelidikan tentang efek jaringan terhadap hasil-hasil internasional kunci, dan kemampuan aktor untuk meningkatkan kekuatan mereka dengan meningkatkan dan memanfaatkan posisi mereka dalam jaringan.

IV.1. Jaringan

Ilmu jaringan yang sekarang dikenal dengan istilah *network science* merupakan penerapan dari teori graf pada bidang matematika. Jaringan dapat dibentuk dengan menggunakan titik (*node* atau *vertex*) dan garis (*link* atau *edge*) (Golbeck, 2013). Dalam jaringan sosial, titik (*node*) dapat digunakan sebagai perwakilan pelaku (aktor) dan garis (*edge*) sebagai perwakilan relasi yang dimiliki oleh dua aktor. Titik dan garis yang digunakan sebagai perwakilan data yang dimodelkan dalam jaringan memiliki beberapa atribut seperti label, bobot, maupun arah (hanya garis).

Garis yang menghubungkan dua titik dapat dibentuk berdasarkan berbagai alasan, salah satunya adalah kesamaan atribut dua titik yang disebut sebagai *homophily*. Jaringan yang

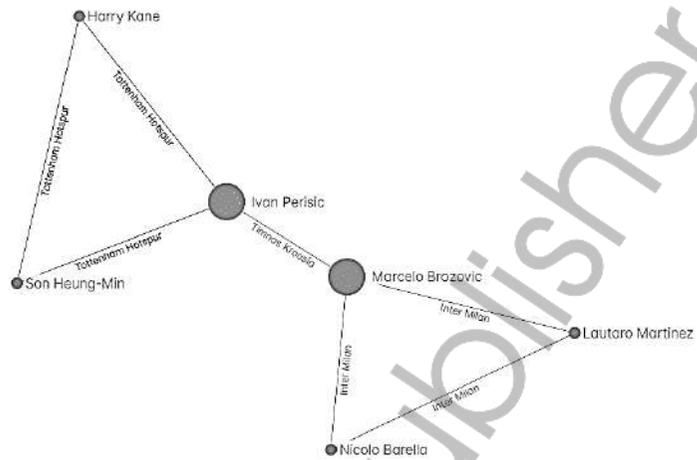
dibentuk berdasarkan kesamaan atribut dari titik dapat dilihat dari beberapa aspek sebagai berikut.

1. Kontak, relasi yang dibentuk berdasarkan kumpulan yang sudah lama bersama;
2. Transaksi, relasi yang dibentuk dari pergaulan dengan orang dari budaya atau daerah yang sama;
3. Sosial, relasi yang dibentuk pada pergaulan orang dengan strata sosial yang sama.



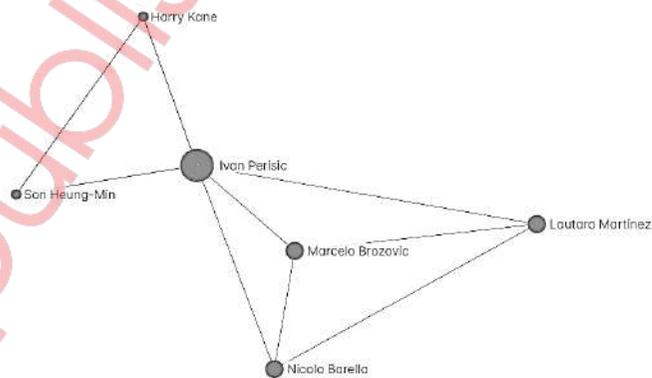
Gambar IV.1. Pemain Sepak Bola Profesional

Penggambaran titik diperlihatkan pada Gambar IV.1 dengan menggunakan pemain sepak bola profesional sebagai contoh aktor, akan tetapi belum ada relasi yang ditetapkan sehingga belum membentuk suatu jaringan. Jika relasi yang ditetapkan adalah kesamaan tim sepak bola, maka jaringan akan terbentuk seperti diperlihatkan pada Gambar IV.2.



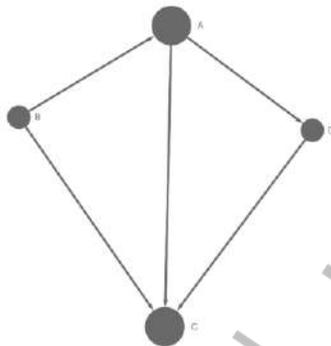
Gambar IV.2. Pemain Sepak Bola dengan Hubungan Tim Sepak Bola yang Sama

Jaringan dapat berubah, jika terdapat perbedaan terhadap relasi yang digunakan sebagai penghubung dua aktor. Misalnya relasi yang digunakan adalah relasi pernah bermain bersama, maka jaringan akan berubah seperti diperlihatkan pada Gambar IV.3.



Gambar IV.3 Pemain Sepak Bola yang Pernah Bermain Bersama

Jaringan berarah disebut juga *directed network* memiliki relasi yang memiliki arah. Karena relasi antara dua titik memiliki arah sehingga dia tidak simetris dan digambarkan dengan garis yang memiliki panah. Titik asal berada pada pangkal panah, sedangkan titik tujuan berada pada ujung panah seperti diperlihatkan pada Gambar IV.4.



Gambar IV.4. Jaringan dengan Relasi Berarah

IV.2. Representasi Jaringan

Representasi jaringan tidak hanya bisa dilakukan dengan menggambarkan titik dan garis, tetapi juga dapat dibuat dalam beberapa bentuk seperti daftar ketetanggaan (*adjacency lists*), matriks ketetanggaan (*adjacency matrix*), hingga dalam bentuk XML (*extensible markup language*) (Golbeck, 2013).

Daftar Ketetanggaan

Daftar ketetanggaan seringkali disebut juga sebagai daftar garis (*edge lists*). Bentuk daftar ketetanggaan dari jaringan pada Gambar IV.3 diperlihatkan pada Tabel IV.1.

Tabel IV.1. Representasi dalam Bentuk Daftar Ketetangaan

Ivan Perisic	Nicolo Barella
Ivan Perisic	Lautaro Martinez
Ivan Perisic	Marcelo Brozovic
Ivan Perisic	Harry Kane
Ivan Perisic	Son Heung-Min
Nicolo Barella	Lautaro Martinez
Nicolo Barella	Marcelo Brozovic
Lautaro Martinez	Marcelo Brozovic
Harry Kane	Son Heung-Min

Jika relasi maupun jaringan yang dibentuk memiliki bobot, maka bobot dapat ditampilkan pada daftar ketetangaan yang diperlihatkan pada Tabel IV.2. Sementara itu representasi jaringan berarah pada Gambar IV.4 diperlihatkan pada Tabel IV.3.

Tabel IV.2. Daftar Ketetangaan dengan Bobot

Ivan Perisic	Nicolo Barella	1
Ivan Perisic	Lautaro Martinez	1
Ivan Perisic	Marcelo Brozovic	1
Ivan Perisic	Harry Kane	1
Ivan Perisic	Son Heung-Min	1
Nicolo Barella	Lautaro Martinez	1
Nicolo Barella	Marcelo Brozovic	1
Lautaro Martinez	Marcelo Brozovic	1
Harry Kane	Son Heung-Min	1

Tabel IV.3. Daftar Ketetangan dari Jaringan Berarah pada Gambar IV.4

A	D	1
A	C	1
B	C	1
B	A	1
D	C	1

Matriks ketetangaan

Bentuk lain yang dapat digunakan untuk merepresentasikan jaringan adalah dengan membentuk matriks ketetangaan (*adjacency matrix*). Matriks ketetangaan menempatkan titik pada baris maupun kolom, kemudian sel yang menjadi pertemuan antara dua titik akan diberi angka sesuai bobot. Matriks berbentuk simetris bagi jaringan tidak berarah. Bagi jaringan tidak berbobot akan diisi nilai 1 jika dua titik memiliki hubungan dan nilai 0 jika tidak.

Tabel IV.4. Matriks Ketetangaan dari Gambar IV.3.

	Marcelo Brozovic	Nicolo Barella	Lautaro Martinez	Ivan Perisic	Hari Kane	Son Heung-Min
Marcelo Brozovic	0	1	1	1	0	0
Nicolo Barella	1	0	1	1	0	0
Lautaro Martinez	1	1	0	1	0	0
Ivan Perisic	1	1	1	0	1	1
Hari Kane	0	0	0	1	0	1
Son Heung-Min	0	0	0	1	1	0

Tabel IV.5. Matriks Ketetangan dari Jaringan Berarah pada Gambar IV.4.

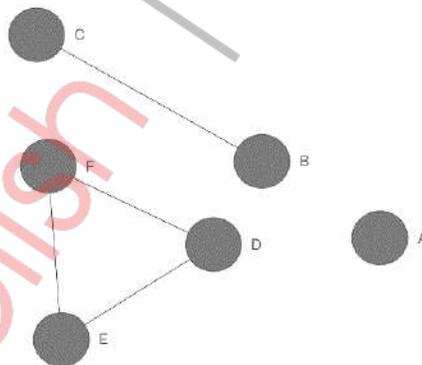
	A	B	C	D
A	0	0	1	1
B	1	0	1	0
C	0	0	0	0
D	0	0	1	0

IV.3. Struktur Jaringan

Selain titik dan garis, terdapat beberapa struktur dasar jaringan yang penting untuk diketahui dan dipahami dalam upaya untuk memahami sebuah jaringan seperti *subnetworks*, *cliques*, *components*, *bridges*, *hubs*, dan *cutpoints*

Subnetworks

Pada dasarnya sebuah *subnetwork* merupakan sebuah jaringan bagian (sub-jaringan) dari sebuah jaringan yang lebih besar (Golbeck, 2013). Jaringan bagian paling sederhana disebut *singleton* karena dia merupakan sebuah titik yang tidak terhubung dengan titik lainnya. Selain *singleton*, terdapat juga *dyad* yang merupakan dua titik yang saling terhubung, tetapi terisolasi dari titik-titik lain dalam jaringan. Sementara itu, jika jaringan bagian itu dibentuk dari tiga titik, maka akan disebut *triad*.



Gambar IV.5. Terdapat Tiga Jaringan Bagian dengan di mana A Merupakan Singleton; B dan C Merupakan Dyad; D, E, dan F Merupakan Triad

Cliques

Ketika semua titik pada suatu kelompok dalam jaringan saling terhubung, maka kelompok itu disebut *clique*. Pada Gambar IV.5, titik D, E, dan F membentuk sebuah *clique*.

Component

Komponen atau *component* merupakan kumpulan titik yang terhubung melalui paling sedikit satu garis dalam jaringan (Eriyanto, 2014). Selama satu titik terhubung dengan jaringan, walaupun hanya melalui satu garis, maka titik tersebut tetap dianggap sebagai bagian dari komponen.

Bridges

Jembatan atau *bridge* merupakan garis yang menghubungkan dua kelompok yang terpisah dalam jaringan (Eriyanto, 2014; Golbeck, 2013). Jika garis yang menghubungkan dua kelompok dihapus atau diputus, maka struktur dan sentralitas dapat berubah.

Hubs

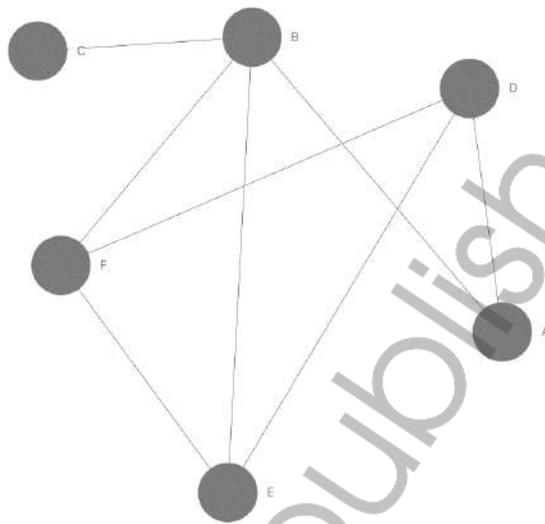
Titik yang memiliki hubungan terbanyak dalam jaringan disebut sebagai *hubs* (Golbeck, 2013).

Cutpoints

Titik yang menjadi perekat dalam jaringan. Tanpa adanya *cutpoints*, jaringan akan terpecah (Eriyanto, 2014). Sebuah *cutpoints* memiliki peran yang mirip dengan *bridge*.

IV.4. Aliran Jaringan

Keterhubungan dalam jaringan membentuk sebuah aliran yang memungkinkan dua buah titik yang tidak terhubung secara langsung tetap bisa terhubung. Terdapat setidaknya tiga bentuk aliran, yaitu *walks*, *paths*, dan *trail*.



Gambar IV.6. Ilustrasi Aliran Jaringan

Walks

Jalan atau *walks* merupakan aliran yang memungkinkan seseorang/ sesuatu menghubungkan semua titik maupun garis dalam jaringan tanpa ada batasan (Eriyanto, 2014). Dengan menggunakan Gambar IV.6 sebagai patokan, jika seorang dari titik C ingin mengunjungi titik A, maka semua kemungkinan jalur yang dapat dilewati adalah *walks* tanpa memandang perulangan titik maupun garis.

Paths

Lintasan atau *paths* dalam jaringan merupakan sebuah rangkaian titik yang dapat dicapai melalui garis yang menghubungkan mereka (Golbeck, 2013). Dalam lintasan, tidak boleh ada pengulangan titik yang dilalui. Misalkan seseorang dari titik C ingin mengunjungi titik A. Terdapat beberapa jalur yang dapat dilalui seperti C-B-A, ini merupakan lintasan (*path*) karena tidak ada pengulangan titik. Bisa juga melalui jalur C-B-F-E-B-A

tetapi ini bukan merupakan lintasan karena melalui melalui titik B secara berulang.

Trails

Jejak atau *trails* memiliki aturan yang mirip dengan lintasan dengan perbedaan terletak pada apa yang tidak boleh diulang. Jika tidak boleh ada titik yang diulang pada lintasan, maka pada jejak tidak boleh ada garis yang diulang (Eriyanto, 2014). Misalkan seseorang dari titik C ingin mengunjungi titik A, maka jalur C-B-F-E-B-A dapat disebut jejak, tetapi jalur C-B-F-E-B-F-D-A tidak bisa disebut jejak karena ada pengulangan garis B-F.

IV.5. Pengukuran dalam Jaringan

IV.5.1. Sentralitas

Sentralitas (*centrality*) merupakan salah satu inti dari analisis jaringan karena merupakan pengukuran untuk mengukur seberapa sentral sebuah titik pada jaringan (Golbeck, 2013). Terdapat beberapa pengukuran sentralitas yang dapat memberikan hasil analisis berbeda pada suatu jaringan berdasarkan sudut pandang maupun konteks analisis yang dilakukan. Disini akan dibahas beberapa pengukuran sentralitas yaitu, sentralitas tingkatan (*degree centrality*), sentralitas kedekatan (*closeness centrality*), sentralitas keperantaraan (*betweenness centrality*) dan sentralitas eigenvektor (*eigenvector centrality*).

Sentralitas Tingkatan

Sentralitas tingkatan (*degree centrality*) merupakan salah satu pengukuran sentralitas yang paling mudah dilakukan. Nilai dari sentralitas tingkatan merupakan jumlah garis yang terhubung pada sebuah titik. Semakin tinggi tingkatan dari satu titik, maka titik tersebut akan dianggap semakin sentral (Golbeck, 2013; Zhang & Luo, 2017). Sentralitas tingkatan dapat dihitung menggunakan rumus berikut (Tahalea *et al.*, 2022).

$$C_D(i) = \sum \frac{d_i}{n-1}$$

di mana i merupakan titik atau aktor yang dihitung sentralitas tingkatannya, n merupakan banyaknya titik atau aktor dalam jaringan, dan d_i merupakan jumlah garis yang terhubung pada aktor i .

Rumus ini sebenarnya cukup sederhana. Untuk menghitung pentingnya seseorang dalam jaringan, kita perlu tahu dua hal. Pertama, berapa banyak orang yang ada dalam jaringan tersebut. Kedua, berapa banyak hubungan yang dimiliki oleh orang yang kita ingin hitung pentingnya.

Misalnya, kita memiliki jaringan teman-teman di media sosial. Ada 100 orang dalam jaringan kita. Lalu, kita ingin tahu seberapa pentingnya seseorang dalam jaringan tersebut, katakanlah orang A. Jika orang A memiliki 10 hubungan dengan orang lain dalam jaringan, maka kita dapat menggunakan rumus ini.

Kita menjumlahkan jumlah hubungan orang A dengan orang lain dalam jaringan, yaitu 10. Kemudian, kita membaginya dengan jumlah orang dalam jaringan kita dikurangi 1 ($100-1=99$). Hasilnya akan memberi kita nilai yang menunjukkan seberapa pentingnya orang A dalam jaringan tersebut.

Mari kita bandingkan dua orang dalam jaringan tersebut, yaitu orang A dan orang B. Kita telah mengetahui bahwa orang A memiliki 10 hubungan dengan orang lain dalam jaringan, sedangkan orang B lebih populer dengan memiliki 50 hubungan.

Kita dapat menggunakan rumus yang telah dijelaskan sebelumnya untuk menghitung pentingnya orang A dan orang B dalam jaringan tersebut. Misalnya, jika jaringan kita terdiri dari 100 orang, kita akan memasukkan nilai "n" sebagai 100.

Untuk orang A, kita menjumlahkan jumlah hubungannya, yaitu 10, dan membaginya dengan $(n-1)$, yang dalam kasus ini

adalah $(100-1) = 99$. Setelah kita menghitung, mungkin kita mendapatkan nilai sentralitas tingkatan untuk orang A sebesar 0,1.

Sementara itu, untuk orang B, dengan 50 hubungan, kita menjumlahkannya dan membaginya dengan $(n-1)$, yang tetap 99. Hasilnya mungkin akan memberikan nilai sentralitas tingkatan yang lebih tinggi untuk orang B, misalnya 0,5.

Dengan melihat nilai sentralitas tingkatan, kita dapat melihat bahwa orang B lebih penting atau lebih populer dalam jaringan tersebut daripada orang A. Meskipun keduanya memiliki pengaruh, orang B memiliki lebih banyak hubungan yang memberikan kepadanya sentralitas tingkatan yang lebih tinggi dalam jaringan.

Sentralitas Kedekatan

Sentralitas kedekatan (*closeness centrality*) menunjukkan kedekatan satu titik ke semua titik lain dalam sebuah jaringan (Tahalea *et al.*, 2022; Zhang & Luo, 2017). Hal ini dapat dilakukan dengan menghitung rata-rata jalur terpendek dari satu titik ke titik lainnya dalam jaringan (Golbeck, 2013), seperti diperlihatkan pada rumus berikut.

$$C_c(i) = \frac{n-1}{\sum d_{i,j}}$$

di mana i merupakan titik atau aktor yang dihitung nilai sentralitas kedekatannya, n merupakan banyaknya titik atau aktor yang ada dalam jaringan, dan $d_{i,j}$ merupakan jarak terpendek dari titik i ke titik j .

Mari kita gunakan contoh jalur pesawat dari beberapa kota di Indonesia untuk menjelaskan konsep ini dengan lebih sederhana. Anggaplah kita memiliki jaringan jalur penerbangan antara kota-kota Jakarta, Bali, Surabaya, Medan, dan Makassar.

Mari kita bandingkan antara Jakarta dan Medan dalam jaringan jalur pesawat di Indonesia. Kita ingin mengetahui seberapa dekat kedua kota ini dengan kota-kota lain dalam jaringan tersebut.

Pertama, kita perlu mengetahui berapa banyak kota yang ada dalam jaringan tersebut, yaitu 5, sehingga nilai "n" adalah 5.

Kemudian, kita mencari jalur terpendek dari Jakarta dan Medan ke setiap kota lain dalam jaringan. Misalnya, jalur terpendek dari Jakarta ke Bali adalah langsung, dari Jakarta ke Surabaya melalui penerbangan transit, dari Jakarta ke Makassar langsung, dan dari Jakarta ke Medan melalui penerbangan transit. Sedangkan untuk Medan, jalur terpendek dari Medan ke Bali melalui penerbangan transit, dari Medan ke Surabaya melalui penerbangan transit, dari Medan ke Makassar langsung, dan dari Medan ke Jakarta melalui penerbangan transit.

Dari sini terlihat bahwa Jakarta memiliki jalur langsung dengan dua kota dan melalui transit dengan dua kota. Sementara Medan hanya satu kota yang bisa langsung sementara tiga yang lain harus transit. Jalur langsung misal kita beri nilai 1, jalur transit diberi nilai 2.

Setelah kita memiliki informasi ini, kita dapat menggunakan rumus yang diberikan sebelumnya untuk menghitung sentralitas kedekatan Jakarta dan Medan. Rumusnya adalah $(n-1)$ dibagi dengan jumlah total jalur terpendek dari Jakarta atau Medan ke semua kota lain dalam jaringan.

Untuk Jakarta, nilai sentralitas *closeness* adalah $4/(1+1+2+2) = 0.66$

Untuk Medan, nilai sentralitas *closeness* adalah $4/(1+2+2+2) = 0.57$

Dengan nilai sentralitas kedekatan ini, kita dapat membandingkan seberapa dekat Jakarta dan Medan dengan kota-kota lain dalam jaringan jalur pesawat. Dalam contoh ini, Medan memiliki nilai sentralitas kedekatan yang sedikit lebih rendah daripada Jakarta, sebab lebih banyak harus transit, menunjukkan

bahwa Medan memiliki keterhubungan yang lebih lemah dengan kota-kota lain dalam jaringan jalur pesawat tersebut.

Sentralitas Keperantaraan

Sentralitas keperantaraan (*betweenness centrality*) mengukur seberapa penting suatu titik terhadap jalur terpendek yang ada pada jaringan (Golbeck, 2013). Sentralitas keperantaraan dapat menunjukkan posisi titik atau aktor sebagai suatu penghubung atau penyalur informasi bagi aktor lain dalam jaringan. Sentralitas keperantaraan dapat dihitung dengan menggunakan rumus berikut.

$$C_B(i) = \sum \frac{g_{j,k}(i)}{g_{j,k}}$$

di mana i merupakan titik yang diukur sentralitas keperantaraannya, $g_{j,k}(i)$ merupakan jalur terpendek dari titik j ke titik k yang melalui titik i , dan $g_{j,k}$ merupakan jalur terpendek dari titik j ke titik k dalam jaringan.

Bayangkan kamu dan teman-temanmu berada di taman bermain yang besar. Kalian semua ingin bermain bersama dan saling berbagi informasi atau pesan satu sama lain. Namun, terkadang tidak mudah bagi semua orang untuk berbicara langsung satu sama lain karena taman bermainnya begitu luas.

Nah, sentralitas keperantaraan adalah cara untuk mengukur seberapa pentingnya peranmu dalam membantu orang lain berkomunikasi dan berbagi informasi. Kamu bisa membayangkan dirimu sebagai "jembatan" atau "pesan" antara teman-temanmu. Jika teman-temanmu ingin saling menyampaikan pesan, mereka mungkin harus melewati kamu.

Semakin banyak pesan atau informasi yang melewati dirimu, semakin tinggi sentralitas keperantaraanmu. Ini berarti

kamu memainkan peran penting dalam menghubungkan orang-orang yang berbeda dan membantu mereka berkomunikasi. Kamu seperti orang yang populer yang diandalkan oleh semua orang untuk menyampaikan pesan.

Jadi, sentralitas keberantaraan mengukur seberapa sering kamu berada di tengah-tengah komunikasi dan seberapa pentingmu sebagai penghubung dalam kelompok. Ini seperti menjadi jembatan atau pesan antara teman-temanmu di taman bermain.

Sentralitas Eigenvektor

Sentralitas eigenvektor (*eigenvector centrality*) merupakan pengukuran yang memperhatikan kepentingan dari titik tetangganya (Golbeck, 2013). Sentralitas ini terkadang digunakan untuk melihat kepopuleran suatu titik. Salah satu contohnya adalah seorang biasa dengan jumlah 100 orang *follower* pada Instagram memiliki nilai sentralitas eigenvektor yang lebih rendah daripada selebgram yang memiliki 100 orang *follower* yang merupakan selebgram juga. Sentralitas eigenvektor dapat dihitung menggunakan rumus berikut.

$$C_E(i) = \frac{1}{\lambda} \sum_{j \in G} a_{i,j}$$

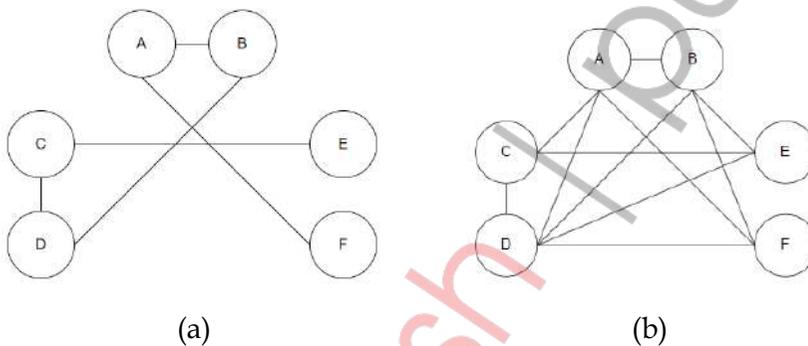
di mana i merupakan titik yang dihitung nilai sentralitasnya, λ adalah konstanta, dan $a_{i,j}$ merupakan matriks ketetanggaan dari i pada jaringan G .

IV.5.2. Densitas

Hubungan satu titik dengan titik lainnya dalam jaringan menunjukkan peran dari titik tersebut dalam jaringan. Untuk memahami titik secara individu dan jaringan secara utuh perlu mempelajari densitas atau kepadatan jaringan.

Menghitung densitas

Densitas menunjukkan bagaimana sebuah jaringan saling terhubung. Secara formal, dapat disebutkan bahwa densitas merupakan perbandingan jumlah garis yang ada dalam jaringan dengan jumlah garis yang mungkin terbentuk (Golbeck, 2013).



Gambar IV.7. Perbandingan Densitas Dua Jaringan

Pada Gambar IV.7. terlihat bahwa jaringan (b) memiliki garis lebih banyak daripada jaringan (a) walaupun keduanya memiliki jumlah titik yang sama, sehingga jaringan (b) memiliki densitas yang lebih tinggi. Adapun densitas jaringan dapat dihitung dengan menggunakan rumus berikut.

$$\text{densitas} = \frac{\text{jumlah garis}}{\text{jumlah garis yang mungkin dibentuk}}$$

Densitas pada jaringan egosentris

Densitas seringkali digunakan untuk membandingkan subjaringan, khususnya jaringan egosentris (*egocentric network*) (Golbeck, 2013). Semakin tinggi nilai densitas pada jaringan egosentris, semakin tinggi tingkat kedekatan antar titik yang ada dalam jaringan. Perhitungan densitas pada jaringan egosentris disebut juga sebagai *local clustering coefficient* (Golbeck, 2013), di mana mempertimbangkan koneksi dari titik dan koneksi yang terbentuk di antara titik-titik dalam jaringan. Perhitungan nilai *local clustering coefficient* dapat menggunakan rumus berikut (Opsahl, 2013).

$$C(i) = \frac{a(i)}{p(i)}$$

di mana i merupakan titik yang akan dicari nilai *local clustering coefficient*-nya, $a(i)$ merupakan jumlah hubungan yang dimiliki oleh titik yang terhubung dengan i , dan $p(i)$ merupakan jumlah hubungan yang mungkin dibentuk oleh titik yang terhubung dengan i .

IV.5.3. Konektivitas

Konektivitas atau *connectivity* merupakan jumlah minimum titik yang perlu dihapus agar membuat suatu jaringan menjadi tidak terhubung (*disconnected*) (Golbeck, 2013). Jaringan (a) pada **Gambar 10** memiliki *connectivity* 1 di mana jika titik C dihapus, maka jaringan akan terputus dan tidak ada jalur yang bisa dibangun ke titik E.

IV.5.4. Sentralisasi

Sentralisasi (*centralization*) merupakan suatu pengukuran yang penting dalam memahami peran suatu titik dalam jaringan. Jika nilai sentralitas satu titik tinggi dan nilai sentralitas titik lain

rendah, maka sentralisasi jaringan tersebut tinggi. Jika nilai sentralitas merata pada semua titik dalam jaringan, maka sentralisasi jaringan tersebut rendah. Dengan kata lain, sentralisasi dihitung dengan melihat jumlah perbedaan sentralitas pada titik paling sentral (tergantung pengukuran sentralitas yang digunakan) dengan titik lainnya dalam jaringan (Freeman, 1978) dalam (Golbeck, 2013).

IV.6. Contoh Penggunaan Analisis Jaringan dalam Riset HI

Analisis jaringan telah menjadi pendekatan yang semakin digunakan dalam penelitian ilmu politik, terutama dalam bidang hubungan internasional. Pendekatan ini memberikan sudut pandang baru dan berbeda dalam memahami struktur, dinamika, dan pengaruh jaringan dalam konteks transnasional, seperti aktivisme transnasional, jaringan teroris, serta perdagangan dan keuangan internasional.

Beberapa contoh:

Jaringan Aktivistis Transnasional (Keck & Sikkink, 1998): Analisis jaringan telah digunakan untuk mempelajari struktur dan dinamika jaringan aktivis transnasional, seperti yang fokus pada hak asasi manusia, isu lingkungan, atau kesehatan global. Para peneliti telah menggunakan analisis jaringan untuk menyelidiki bagaimana jaringan ini terbentuk, bagaimana mereka memengaruhi politik internasional, dan bagaimana mereka mempromosikan tata kelola global. Sebagai contoh, sebuah studi menggunakan analisis jaringan untuk memeriksa struktur jaringan hak asasi manusia global dan menemukan bahwa jaringan tersebut sangat terpusat di sekitar beberapa organisasi kunci, yang memiliki pengaruh yang tidak sebanding pada agenda dan aktivitas jaringan.

Jaringan Teroris (JORDAN & HORSBURGH, 2005): Analisis jaringan juga telah digunakan untuk mempelajari struktur dan dinamika jaringan teroris, seperti Al Qaeda atau ISIS. Para peneliti

telah menggunakan analisis jaringan untuk memahami bagaimana jaringan ini terbentuk, bagaimana mereka merekrut anggota, dan bagaimana mereka merencanakan dan melaksanakan serangan. Analisis jaringan juga telah digunakan untuk mengembangkan strategi untuk melawan terorisme, seperti mengidentifikasi simpul kunci dalam jaringan yang dapat ditargetkan untuk gangguan. Sebagai contoh, sebuah studi menggunakan analisis jaringan untuk memeriksa struktur jaringan Al Qaeda dan menemukan bahwa jaringan tersebut sangat terdesentralisasi, dengan banyak sel kecil yang beroperasi secara independen tetapi terhubung melalui beberapa individu kunci.

Perdagangan dan Keuangan Internasional (Kuncoro & Sari, 2022): Analisis jaringan telah digunakan untuk mempelajari struktur dan dinamika jaringan ekonomi global, seperti yang terlibat dalam perdagangan internasional atau keuangan. Peneliti menyelidiki dinamika perdagangan antara negara-negara Indo-Pasifik dan Eropa menggunakan analisis jaringan. Tujuannya adalah meningkatkan pemahaman kita tentang saling ketergantungan antara wilayah-wilayah ini dengan mengidentifikasi pola kedekatan dan interaksi di antara negara-negara tersebut. Dokumen ini menyajikan gambar dan tabel yang menunjukkan nilai perdagangan total dan kesamaan dalam impor dan ekspor. Selain itu, dokumen ini membahas signifikansi negara-negara Indo-Pasifik dalam perdagangan global dan menjelaskan penggunaan teori jaringan dalam pemetaan hubungan perdagangan yang rumit. Selain itu, dokumen ini menguraikan metodologi yang digunakan dalam penelitian ini, yang merupakan pendekatan metode campuran yang menggabungkan metode kuantitatif dan kualitatif. Analisis Jaringan Sosial (SNA) berfungsi sebagai alat analisis utama untuk memetakan hubungan perdagangan antara anggota Indo-Pacific Economic Forum (IPEF)-plus dan Uni Eropa, yang memungkinkan gambaran menyeluruh

dan identifikasi "aktor sentral" menggunakan berbagai parameter. Dokumen ini juga mencakup representasi visual hubungan perdagangan melalui gambar dan tabel yang menampilkan hasil analisis jaringan. Secara keseluruhan, dokumen komprehensif ini memberikan wawasan berharga tentang hubungan perdagangan yang kompleks antara negara-negara Indo-Pasifik dan Eropa.

Referensi

- Borgatti, S. P., & Halgin, D. S. (2011). On Network Theory. *Organization Science*, 22(5), 1168-1181. <https://doi.org/10.1287/orsc.1100.0641>
- Eriyanto. (2014). *Analisis jaringan komunikasi: strategi baru dalam penelitian ilmu komunikasi dan ilmu sosial lainnya: dilengkapi tutorial penggunaan UCINET-NetDraw*. PRENADAMEDIA GROUP.
- Freeman, L. C. (1978). Centrality in social networks conceptual clarification. *Social Networks*, 1(3), 215-239. [https://doi.org/https://doi.org/10.1016/0378-8733\(78\)90021-7](https://doi.org/https://doi.org/10.1016/0378-8733(78)90021-7)
- Golbeck, Jennifer. (2013). *Analyzing the social web*. Morgan Kaufmann.
- Granovetter, M. (1976). Network Sampling: Some First Steps. *American Journal of Sociology*, 81(6), 1287-1303. <https://doi.org/10.1086/226224>
- Knoke, D., & Kuklinski, J. H. (1982). *Network Analysis*. SAGE Publishing.
- Opsahl, T. (2013). Triadic closure in two-mode networks: Redefining the global and local clustering coefficients. *Social Networks*, 35(2), 159-167. <https://doi.org/https://doi.org/10.1016/j.socnet.2011.07.001>
- Scott, J. (2017). *Social Network Analysis* (4th ed.). SAGE Publications, Inc. <https://doi.org/https://doi.org/10.4135/9781529716597>

- Tahalea, S., Salouw, E., & Wibowo, A. (2022). TOURISM ON INSTAGRAM: A SOCIAL NETWORK ANALYSIS. *BAREKENG: Jurnal Ilmu Matematika Dan Terapan*, 16(1), 199–206. <https://doi.org/10.30598/barekengvol16iss1pp197-204>
- Varda, D. M., Forgette, R., Banks, D., & Contractor, N. (2009). Social network methodology in the study of disasters: Issues and insights prompted by post-katrina research. In *Population Research and Policy Review* (Vol. 28, Issue 1, pp. 11–29). <https://doi.org/10.1007/s11113-008-9110-9>
- Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications*. Cambridge university press.
- Zhang, J., & Luo, Y. (2017). Degree Centrality, Betweenness Centrality, and Closeness Centrality in Social Network. 300–303. <https://doi.org/10.2991/MSAM-17.2017.68>

■

Bab V

Aplikasi Metode Komputasional

Set by Set

Pada bab ini, kita akan membahas penggunaan metode komputasional dalam penelitian Hubungan Internasional (HI). Kita akan fokus pada dua metode utama, yaitu analisis sentimen dan analisis jaringan. Metode ini dapat memberikan wawasan berharga dalam memahami dinamika HI dengan pendekatan yang lebih ilmiah. Pertama, mari kita lihat analisis sentimen. Metode ini----- memungkinkan kita untuk menganalisis opini, sikap, dan perasaan yang terkandung dalam teks, seperti dokumen kebijakan, laporan berita, atau konten media sosial yang berkaitan dengan isu-isu internasional. Dengan langkah-langkah yang terperinci, kita dapat memahami bagaimana menerapkan analisis sentimen dalam konteks HI. Misalnya, kita dapat mengidentifikasi sentimen positif, negatif, atau netral terhadap kebijakan luar negeri suatu negara atau tanggapan publik terhadap peristiwa internasional tertentu. Melalui contoh-contoh konkret, pembaca dapat melihat bagaimana analisis sentimen dapat memberikan pemahaman yang lebih mendalam tentang persepsi dan reaksi terhadap isu-isu HI.

Selanjutnya, kita akan membahas analisis jaringan. Metode ini membantu kita memetakan hubungan antara negara, organisasi internasional, atau aktor lain dalam sistem HI. Dengan menggunakan alat analisis jaringan, kita dapat memahami struktur, kekuatan koneksi, dan interaksi antaraktor dalam HI. Penjelasan

langkah demi langkah akan memberikan pemahaman yang lebih rinci tentang bagaimana menerapkan analisis jaringan dalam konteks HI. Misalnya, kita dapat mengidentifikasi negara-negara yang memiliki hubungan yang kuat dalam kerja sama ekonomi atau kerja sama keamanan. Contoh-contoh konkret akan memberikan ilustrasi tentang bagaimana analisis jaringan dapat memberikan wawasan yang berharga dalam memahami dinamika HI. Dengan memberikan penjelasan yang lebih rinci dan menggunakan bahasa yang lebih alamiah, diharapkan pembaca dapat dengan mudah memahami penggunaan metode komputasional seperti analisis sentimen dan analisis jaringan dalam penelitian Hubungan Internasional.

V.1. Analisis Sentimen

Penelitian ini mengkaji sentimen pengguna Twitter terkait dengan repatriasi lebih dari seribu individu Indonesia yang sebelumnya terkait dengan Negara Islam dan tinggal di kamp-kamp di Suriah. Melalui analisis sentimen, analisis jaringan, dan penambangan teks, penelitian ini bertujuan untuk mengidentifikasi kekhawatiran yang umum diungkapkan oleh pengguna Twitter terkait repatriasi mantan anggota Negara Islam, mengkaji pengaruh akun pemerintah dan non-pemerintah dalam perdebatan di Twitter, dan menyoroti implikasi keamanan dan masalah kemanusiaan yang harus dipertimbangkan oleh pembuat kebijakan dalam pengambilan keputusan tentang repatriasi. Hasil penelitian menunjukkan bahwa repatriasi individu-individu ini menimbulkan perhatian dan kekhawatiran yang signifikan di kalangan publik. Oleh karena itu, pembuat kebijakan harus mempertimbangkan implikasi keamanan dan masalah kemanusiaan yang terlibat dalam pengambilan keputusan semacam itu.

Metode pengumpulan data yang umum digunakan dalam analisis media sosial terkait dengan topik ini adalah dengan

mengumpulkan *tweet* berdasarkan kata kunci yang berkaitan dengan ISIS. Pertanyaan pencarian yang berisi kata kunci relevan seperti "ISIS" dan "Islamic State" digunakan dalam alat atau *platform* analisis media sosial untuk mengumpulkan *tweet* yang mencakup kata kunci tersebut. Pemilihan kata kunci yang tepat sangat penting agar *tweet* yang terkumpul benar-benar relevan dengan topik penelitian.

Pengumpulan *tweet* dilakukan dalam periode tertentu untuk memfokuskan analisis pada *tweet* yang terkait dengan perkembangan terbaru dan isu terkini seputar ISIS. Metode pengumpulan *tweet* berdasarkan kata kunci ini memungkinkan peneliti untuk mendapatkan informasi langsung dari pengguna media sosial terkait dengan isu ini. Dengan menggunakan alat dan teknik analisis yang sesuai, peneliti dapat memproses, menganalisis, dan menginterpretasikan *tweet-tweet* tersebut untuk memperoleh pemahaman yang lebih dalam tentang isu ISIS dan perubahan opini masyarakat terkait dengan kelompok ini.

Dalam analisis sentimen, penelitian ini menggunakan pendekatan berbasis leksikon, di mana setiap kata dalam leksikon atau kamus memiliki skor sentimen yang telah ditentukan sebelumnya. Skor sentimen ini umumnya merupakan nilai numerik yang menunjukkan tingkat positivitas atau negatitivitas yang terkait dengan kata tersebut. Misalnya, kata "bahagia" mungkin memiliki skor positif +1, sementara kata "sedih" mungkin memiliki skor negatif -1.

Untuk menganalisis sentimen sebuah *tweet*, penelitian ini pertama-tama memecah *tweet* tersebut menjadi kata-kata individu dan mencari skor sentimen untuk setiap kata dalam leksikon. Skor sentimen untuk *tweet* tersebut kemudian dihitung sebagai rata-rata skor sentimen dari semua kata dalam *tweet* tersebut. Jika skor sentimen positif, *tweet* tersebut diklasifikasikan sebagai memiliki

sentimen positif, sedangkan jika skor sentimen negatif, *tweet* tersebut diklasifikasikan sebagai memiliki sentimen negatif.

Selanjutnya, penelitian ini memisahkan *tweet-tweet* ke dalam dua kelompok berdasarkan skor sentimen mereka: kelompok dengan skor positif yang lebih tinggi daripada skor negatif, dan kelompok dengan skor negatif yang lebih tinggi. *Tweet-tweet* dalam setiap kelompok kemudian dianalisis untuk menemukan narasi-narasi yang ada. Dalam analisis sentimen positif terkait dengan isu repatriasi mantan anggota Negara Islam di Indonesia, ditemukan beberapa kata-kata yang membuat orang merasa positif, seperti Indonesia, Pemerintah, Warga Negara Indonesia, Mantan, Anak, Presiden, Repatriasi, dan lain-lain. Hal ini menunjukkan bahwa beberapa orang di Twitter menyampaikan pendapat positif tentang isu repatriasi mantan anggota Negara Islam dari Indonesia.

Namun, ketika melihat lebih dalam pada data sentimen positif yang dikumpulkan, ditemukan bahwa pendapat negatif lebih dominan. Hal ini berarti meskipun ada beberapa orang di Twitter yang merasa positif tentang isu repatriasi mantan anggota Negara Islam, sebagian besar percakapan publik di Twitter tentang isu ini cenderung negatif. Ditemukan bahwa pendapat negatif yang diungkapkan oleh pengguna Twitter didominasi oleh kekhawatiran tentang keamanan dan terorisme. Mereka melihat mantan anggota Negara Islam sebagai ancaman radikalisme bagi masyarakat, dan khawatir bahwa mereka yang kembali ke Indonesia dapat menyebarkan radikalisme dan mengancam keamanan masyarakat. Hal ini menunjukkan bahwa percakapan publik di Twitter tentang isu repatriasi mantan anggota Negara Islam dari Indonesia lebih banyak dipengaruhi oleh kekhawatiran tentang keamanan dan terorisme.

Studi mengenai analisis sentimen menggunakan data Twitter dalam konteks repatriasi mantan anggota Negara Islam dari Indonesia memberikan wawasan berharga tentang pendapat

dan kekhawatiran publik terkait isu ini. Dengan menganalisis wacana publik di Twitter, studi ini mengungkap perhatian dan kekhawatiran yang signifikan yang diungkapkan oleh pengguna terkait proses repatriasi. Temuan tersebut menyoroti dominasi sentimen negatif yang terkait dengan keamanan dan terorisme, menunjukkan pentingnya mempertimbangkan faktor-faktor ini saat membuat keputusan tentang repatriasi.

Namun, penting untuk mengakui keterbatasan studi ini. Ketergantungan pada data Twitter mungkin tidak sepenuhnya mencerminkan sentimen publik secara luas, karena hanya mencakup pandangan dari *subset* tertentu dalam populasi. Selain itu, studi ini berfokus pada analisis sentimen tanpa mengeksplorasi alasan mendasar di balik sentimen yang diungkapkan. Penelitian selanjutnya dapat mengatasi keterbatasan ini dengan menggabungkan metode kualitatif dan mengeksplorasi sentimen di berbagai *platform* media sosial.

Studi ini memberikan kontribusi pada pengetahuan yang ada dengan memberikan pemahaman yang lebih mendalam tentang wacana publik seputar repatriasi dan menekankan perlunya mempertimbangkan implikasi keamanan dan kemanusiaan dalam proses pengambilan keputusan. Para pembuat kebijakan harus memperhatikan kekhawatiran yang diungkapkan oleh publik, terutama terkait keamanan dan terorisme, ketika merumuskan kebijakan repatriasi. Komunikasi yang terbuka dan transparan dengan publik, termasuk melibatkan aktif di *platform* media sosial, dapat membantu mengatasi kekhawatiran, mengklarifikasi kesalahpahaman, dan memperkuat pemahaman yang lebih baik tentang proses repatriasi.

V.2. Analisis Jaringan

Sebelum analisis kerangka dapat dilakukan, data dari Twitter harus dikumpulkan dan diproses. Data akan diekstrak dari Twitter menggunakan aplikasi RStudio dengan paket *academictwitterR* (Barrie & Ho, 2021). Sebanyak mungkin *tweet* akan dikumpulkan menggunakan dua filter, yaitu kata kunci dan negara. Untuk kata kunci, akan digunakan "ISIS", "Daesh", "IS", "Islamic State of Iraq and Syria", "Islamic State of Iraq and the Levant", "ISIL", "#isis", "#is", "#islamic_state", "#Dawla", "#Baqiyah", "pengembalian ISIS", "hak asasi manusia ISIS", dan "Da'ish". Versi lokal kata-kata tersebut juga digunakan untuk memastikan pengumpulan data yang komprehensif. *Tweet* yang mengandung salah satu kata kunci tersebut akan diunduh ke dalam basis data. Filter kedua adalah filter negara, karena penelitian ini bertujuan untuk membandingkan kerangka pemikiran dari negara-negara yang berbeda. Oleh karena itu, *tweet-tweet* spesifik yang dikirim dari negara-negara ini akan dikumpulkan. Negara-negara yang akan menjadi sumber *tweet* adalah:

- *Euro-Anglosphere*: Amerika Serikat, Inggris, Belanda, Jerman, Prancis, Belgia, Austria, Australia.
- Asia: Tiongkok, Afganistan, Pakistan, Indonesia, Malaysia, Filipina.
- Timur Tengah dan Afrika Utara: Arab Saudi, Turki, Tunisia, Mesir, Yordania, Maroko, Suriah, Irak.

Negara-negara ini dipilih karena mereka mencakup jumlah individu yang signifikan yang bergabung dengan ISIS. *Tweet-tweet* yang terkumpul akan dikelompokkan menjadi tiga kelompok seperti yang dijelaskan di atas, berdasarkan kesamaan geografis dan/atau sosio-kultural. Seluruh *tweet* kemudian akan menjalani proses pembersihan. Proses pembersihan ini sangat penting karena akurasi langkah-langkah selanjutnya bergantung pada kebersihan data. Proses pembersihan melibatkan beberapa tahap, termasuk

penghapusan tanda baca, angka, tautan, penyebutan karakter ("@"), dan emotikon. *Tweet-tweet* yang tidak dalam bahasa Inggris akan diterjemahkan ke dalam bahasa Inggris menggunakan paket `googleLanguageR` (Edmondson, 2020). Terjemahan ini menggunakan Google Cloud Translator, yang dikenal karena akurasi terjemahannya yang tinggi. Terjemahan adalah langkah penting karena penghapusan *stopwords* akan dilakukan menggunakan metode yang dirancang untuk *tweet-tweet* berbahasa Inggris. Langkah selanjutnya adalah menghapus *stopwords*, yaitu kata-kata yang sering digunakan tetapi tidak memiliki makna khusus, seperti kata sambung. Sumber *stopwords* berasal dari *database* Natural Language Tool Kits yang dikembangkan oleh Bird (Bird, 2021). Langkah ini membantu memfokuskan analisis pada konten yang bermakna dan relevan.

Untuk memperhalus pengelompokan *tweet*, dilakukan analisis jaringan. Analisis jaringan memfasilitasi pemetaan interaksi pengguna, memungkinkan identifikasi pola dan struktur yang muncul dalam jaringan, dan memberikan wawasan berharga tentang dinamika dan keterhubungan pengguna (Otte & Rousseau, 2016). Untuk penelitian khusus ini, paket "igraph" dalam R digunakan (Csardi & Nepusz, 2006). Analisis jaringan digunakan untuk mengidentifikasi tokoh pendapat kunci (key opinion leaders/KOLs) dalam setiap kelompok negara. KOLs adalah pengguna Twitter yang memiliki pengaruh signifikan dalam percakapan, karena sering kali mereka menerima penyebutan (mention) dan *tweet* mereka sering di-*retweet*, dijawab, dan dikutip. Dalam setiap kelompok negara, ditentukan 10 KOLs dengan tingkat *in-degree* tertinggi. *In-degree* mengacu pada jumlah koneksi masuk dalam bentuk penyebutan, jawaban, kutipan, dan *retweet* yang diterima oleh pengguna dalam jaringan. Metrik ini membantu mengidentifikasi pengguna yang sangat aktif dan berpengaruh dalam komunitas masing-masing. Dari 10 KOLs yang

teridentifikasi, hanya mereka yang terafiliasi dengan lembaga media utama yang dipilih. Kriteria pemilihan ini sesuai dengan tujuan penelitian untuk memeriksa kerangka media. Dengan berfokus pada lembaga media utama dalam 10 KOLs teratas, kami memastikan bahwa analisis kami berpusat pada sumber media yang sangat terlibat dan *tweet* mereka tersebar luas, sehingga memiliki pengaruh lebih besar dalam membentuk opini publik.

Untuk setiap akun media yang teridentifikasi, dilakukan analisis jaringan kata untuk memvisualisasikan hubungan antara kata-kata yang digunakan dalam *tweet-tweet* tersebut. Jaringan kata menggambarkan kata-kata sebagai simpul dan keterjadian bersama atau kedekatan antara kata-kata sebagai tepi. Analisis sentralitas, khususnya sentralitas *in-degree* (mirip dengan yang digunakan untuk mengidentifikasi KOL sebelum langkah ini), kemudian dilakukan pada jaringan kata untuk mengidentifikasi kata-kata paling berpengaruh yang terkait dengan fenomena ISIS di Twitter. Analisis ini bertujuan untuk mengungkap kata-kata yang memiliki jumlah koneksi terbanyak dengan kata-kata lain dalam jaringan, menunjukkan signifikansi mereka dalam membentuk wacana.

Kata-kata berpengaruh kemudian diinterpretasikan berdasarkan asosiasi kontekstual mereka dengan kata-kata berikutnya dan sebelumnya yang paling sering dalam jaringan kata mereka masing-masing. Interpretasi ini memberikan wawasan tentang bagaimana kata-kata berpengaruh ini di-*framing* dan dihubungkan dengan konsep-konsep lain dalam konteks terbatas data Twitter. Selanjutnya, dilakukan analisis kerangka (*framing*) pada interpretasi kata-kata paling berpengaruh dan kata-kata yang berdekatan dengannya. Para peneliti mengkaji bagaimana kata-kata berpengaruh ini di-*framing* dan unsur-unsur *framing* yang hadir dalam konteks sekitarnya. Analisis ini berfokus pada fungsi definisi masalah dan diagnosis penyebab *framing*, karena kedua fungsi ini

dapat dieksplorasi dengan lebih baik dalam batasan karakteristik data Twitter.

Langkah terakhir melibatkan perbandingan hasil analisis *framing* dari masing-masing kelompok (*Euro-Anglosphere*, Asia, dan Timur Tengah). Analisis perbandingan ini bertujuan untuk mengidentifikasi variasi pola *framing* di berbagai negara. Para peneliti mengkaji prevalensi dan penekanan unsur *framing* definisi masalah dan diagnosis penyebab dalam setiap kelompok, dan menganalisis bagaimana pola *framing* ini berbeda di negara-negara yang berbeda.

Penemuan dari studi ini memperluas pemahaman kita tentang representasi media dan *framing* dalam konteks ancaman keamanan internasional seperti ISIS. Sebelumnya, banyak diskusi ilmiah tentang *framing* media berfokus pada masalah nasional atau regional, tetapi penelitian ini menekankan pengaruh konteks geografis dan politik pada narasi media tentang ancaman global. Perspektif baru ini mendorong para sarjana untuk mempertimbangkan efek situasi geopolitik dan sikap budaya dalam membentuk diskursus media, menekankan kompleksitas dan sifat multifaset dari peristiwa internasional sebagaimana digambarkan oleh media.

Studi ini menawarkan tambahan signifikan ke bidang studi media, studi keamanan global, dan analisis sosio-politik. Analisis komparatif di tiga wilayah yang berbeda menunjukkan bagaimana ancaman global dilokalkan dalam diskursus media dan bagaimana perspektif lokal membentuk pemahaman publik tentang ancaman semacam itu. Dengan demikian, ia mengungkapkan bagaimana media, sebagai institusi sosial, berinteraksi dengan lingkungan politik dan budaya untuk membentuk diskursus publik. Penemuan ini menambah lapisan vital pemahaman pada persepsi kita terhadap peristiwa global, menunjukkan bahwa interpretasi

peristiwa-peristiwa ini sering disesuaikan untuk sejalan dengan perspektif, kepentingan, dan realitas sosio-politik regional.

Studi ini, meskipun penuh wawasan, dibatasi oleh fokusnya pada beberapa media outlet terpilih dari masing-masing wilayah, yang mungkin tidak sepenuhnya mewakili keragaman perspektif media di dalam masing-masing wilayah. Seleksi ini dapat secara tidak sengaja mempengaruhi kesimpulan dan membatasi generalisabilitas penemuan. Selain itu, studi ini sangat bergantung pada interpretasi data teks. Meskipun setiap upaya telah dilakukan untuk memastikan analisis objektif dan teliti, ada subjektivitas yang melekat dalam menginterpretasikan data teks yang mungkin mempengaruhi hasil.

Meskipun dengan keterbatasan yang terkait dengan pemilihan data dan potensi interpretasi subjektif, studi ini memberikan kontribusi signifikan pada pengetahuan yang ada. Ini memberikan analisis terperinci tentang bagaimana narasi media seputar ancaman global dibangun dan bagaimana mereka mungkin berbeda berdasarkan faktor regional. Perbandingan terperinci di antara wilayah, meskipun jumlah media outlet terbatas, menawarkan wawasan berharga tentang interaksi kompleks media, masyarakat, dan geopolitik, sehingga memperkaya pemahaman kita tentang peran media dalam membentuk diskursus publik seputar ancaman keamanan internasional.

Penelitian di masa depan harus bertujuan untuk memperluas cakupan studi ini, baik dalam hal outlet media yang dipertimbangkan dan bentuk media yang dianalisis. Melibatkan berbagai media outlet dari setiap wilayah akan memungkinkan pemahaman yang lebih komprehensif dan nuansa tentang *framing* media regional. Selain itu, menganalisis berbagai bentuk media seperti media sosial, berita televisi, dan bahkan *podcast* dapat menawarkan wawasan berharga tentang bagaimana berbagai bentuk media mempengaruhi persepsi publik. Akan juga berguna

untuk mempelajari penerimaan dan persepsi penonton terhadap narasi media ini untuk mengukur dampak mereka terhadap opini publik dan sikap terhadap ancaman global seperti ISIS.

Berdasarkan penemuan dari studi ini, pembuat kebijakan dan praktisi harus bekerja untuk mempromosikan literasi media di kalangan publik. Mengingat peran penting media dalam membentuk opini dan diskursus publik, penting bagi audiens untuk dapat menganalisis konten media secara kritis, memahami bias inherennya, dan membedakan antara pelaporan objektif dan wacana beropini. Bagi media outlet, penemuan ini menekankan kebutuhan untuk jurnalisme yang bertanggung jawab dan pelaporan seimbang, terutama ketika berhadapan dengan topik sensitif seperti terorisme. Dalam hal kebijakan, studi ini menyoroti pentingnya mempertimbangkan konteks dan persepsi regional dalam merumuskan strategi kontra-terorisme. Pendekatan '*one-size-fits-all*' mungkin tidak efektif mengingat narasi dan perspektif yang berbeda terhadap ISIS di berbagai wilayah, dan strategi harus disesuaikan untuk memperhitungkan perbedaan ini.

Profil Penulis



Dr. Iva Rachmawati

merupakan dosen Ilmu Hubungan Internasional UPN Veteran Yogyakarta. Dr. Iva fokus mengkaji diplomasi dalam Hubungan Internasional. Penelitiannya juga mencakup studi diplomasi dan studi perbatasan. Saat ini, dia juga melakukan riset-riset terkait kontribusi metode komputasional dalam Hubungan Internasional.

Hestutomo restu Kuncoro, M.A.

merupakan dosen Ilmu Hubungan Internasional UPN Veteran Yogyakarta. Sebagai lulusan Ilmu Politik The University of Manchester, dia aktif terlibat sejumlah penelitian tentang konsolidasi demokrasi dan demokratisasi. Saat ini, dia fokus mengkaji metode komputasional dan sosial media dalam isu Hubungan Internasional.



Dyah Lupita Sari, M.Si

merupakan dosen Ilmu Hubungan Internasional UPN Veteran Yogyakarta. Menyelesaikan studi magisternya di Universitas Indonesia dan mengkaji disinformasi sebagai salah satu ancaman kedaulatan negara. Saat ini dia aktif mengkaji peran sosial media dan pengaruh media dalam berbagai isu keamanan internasional.

Sylvert Prian Tahalea, M.Cs.

merupakan dosen Informatika UPN Veteran Yogyakarta dengan konsentrasi sains data dan rekayasa perangkat lunak. Menyelesaikan studi magister dengan penelitian terkait analisis jaringan kriminal berkelompok. Saat ini aktif terlibat dalam penelitian implementasi analisis jaringan pada berbagai bidang.

