

Visualisasi Data Tweet di Sektor Pendidikan Tinggi Pada Saat Masa Pandemi

Djoko C. U. Lieharyani^{1,*}, Rita Ambarwati²

¹Jurusan Teknik Komputer & Informatika, Politeknik Negeri Bandung, Bandung, Indonesia

²Fakultas Bisnis Hukum dan Ilmu Sosial, Universitas Muhammadiyah Sidoarjo, Sidoarjo, Indonesia

Email: ^{1,*}djoko.c.utomo@polban.ac.id, ²ritaambarwati@umsida.ac.id

Email Penulis Korespondensi: djoko.c.utomo@polban.ac.id

Submitted: 28/04/2022; Accepted: 24/05/2022; Published: 30/06/2022

Abstrak—Masa pandemi Covid-19 memberikan dampak ke berbagai sektor, salah satunya adalah pendidikan tinggi. Beragam opini masyarakat di media sosial tentang pendidikan tinggi selama masa pandemi bermunculan dengan berbagai macam fokus diskusi. Tujuan penelitian ini ada dua, yang pertama adalah untuk membandingkan hasil analisis semantik data menggunakan stemming dan tanpa menggunakan stemming. Kedua untuk mengetahui fokus pembahasan dari para pengguna twitter terkait dengan pendidikan tinggi. Penelitian ini dilakukan dengan mengambil data tweet selama setahun masa pandemi dari maret 2020 hingga maret 2021. Data tersebut kemudian dilakukan preprocessing dengan gabungan metode agar secara maksimal tidak ada data noise. Tahapan terakhir kemudian data tersebut dianalisis dengan menggunakan teori informasi, frasa bigram, analisis hubungan semantik antar kata dan divisualisasikan dengan menggunakan gephi. Hasil penelitian ini memberikan penjelasan rinci alasan metode stemming tidak dapat langsung dilakukan untuk preprocessing data yang membutuhkan pemahaman semantik. Visualisasi data juga menunjukkan pengguna twitter yang membahas terkait dengan pendidikan tinggi lebih berfokus ke arah kehidupan kampus dan penerimaan mahasiswa baru. Penelitian ini melakukan evaluasi pada tahapan stemming karena hasil menunjukkan bahwa dengan menggunakan stemming akan menghilangkan semantik antar kata dan memvisualisasikan hubungan semantik pada tweet yang berkaitan dengan pendidikan tinggi..

Kata Kunci: Covid-19; Preprocessing; Semantik; Pendidikan Tinggi; Media Sosial

Abstract—The Covid-19 pandemic has had an impact on many sectors, including higher education. Various public opinions about higher education have emerged on social media during the pandemic, with varying foci of discussion. The purpose of this study is twofold. The first is to compare the results of semantic data analysis with and without stemming. Second, to determine the topic of discussion among Twitter users about higher education. This study was carried out by collecting tweet data for a year during the pandemic, from March 2020 to March 2021. The data then pre-processed using a combination of methods to ensure that there was minimal noise data. The analysis is performed using information theory, bigram phrases, semantic relationship analysis between words, and visualized using gephi in the final stage. The findings of this study explain in detail why the stemming method cannot be used directly for preprocessing data that requires semantic understanding. Data visualization also reveals that Twitter users who talk about higher education are more concerned with campus life and new student admissions. This research evaluates the stemming process because the results show that using stemming eliminates semantic meaning between words. This research also visualizes semantic relationships in tweets about higher education.

Keywords: Covid19; Preprocessing; Semantics; Higher Education; Social Media

1. PENDAHULUAN

Pandemi Covid-19 memberikan dampak pada Indonesia, dimana kasus yang pertama kali terkonfirmasi terjadi di 2 maret 2020, dimana indonesia dianggap terlambat dalam mengkonfirmasi kasus positif ini [1]. Dampak yang sangat signifikan terjadi pada berbagai sektor terutama pada bidang pendidikan karena kewajiban diterapkannya kebijakan sosial distancing, PSBB (pembatasan sosial berskala besar), physical distancing[2]. Pemerintah menerapkan kebijakan untuk meliburkan siswa, mahasiswa dan pekerja dan menggantinya dengan WFH (Work From Home), SFH (Study From Home) [1]. Hal ini membuat perguruan tinggi diminta untuk menyiapkan kebijakan untuk pembelajaran secara online. Berbagai opini pun muncul di media sosial terkait kebijakan yang diambil oleh perguruan tinggi dalam menghadapi pandemi.

Pendidikan di Indonesia sendiri terbagi menjadi tiga jenis pendidikan tinggi yaitu pendidikan akademik (sarjana, magister, dan doktor), pendidikan profesi/spesialis, dan pendidikan vokasi (diploma/sarjana terapan) [3]. Pendidikan akademik merupakan program sarjana untuk penguasaan serta pengembangan cabang ilmu pengetahuan dan teknologi, pendidikan akademik ini secara umum dijalankan oleh universitas dan institut. Pendidikan vokasi adalah pendidikan tinggi untuk diploma dengan tujuan untuk menyiapkan generasi muda dalam pekerjaan yang membutuhkan keahlian terapan tertentu, pendidikan vokasi ini dijalankan oleh politeknik dan sekolah tinggi. Pendidikan profesi merupakan pendidikan tinggi yang dilakukan setelah menyelesaikan program pendidikan sarjana untuk menyiapkan peserta didik atau mahasiswa agar memiliki pekerjaan dengan persyaratan keahlian khusus seperti dokter. Pendidikan tinggi adalah pusat pengembangan dari berbagai macam ilmu pengetahuan dan teknologi yang ditargetkan untuk dapat meningkatkan, memajukan dan mempercepat kegiatan pembangunan dan pengembangan nasional [4]. Pendidikan tinggi perlu untuk terus berkembang mengikuti zaman, tetapi dalam perkembangannya terdapat banyak masalah seperti penyalarsan strategi bisnis [5].

Lulusan dan masyarakat adalah stakeholder yang berkepentingan terhadap hasil kontribusi dari perguruan tinggi [6], karena itu perlu ada formulasi strategi yang tepat sesuai kebutuhan dari stakeholder dan perkembangan zaman. Media sosial adalah salah satu bentuk strategi bisnis yang tepat untuk menyampaikan hasil produk pada masa

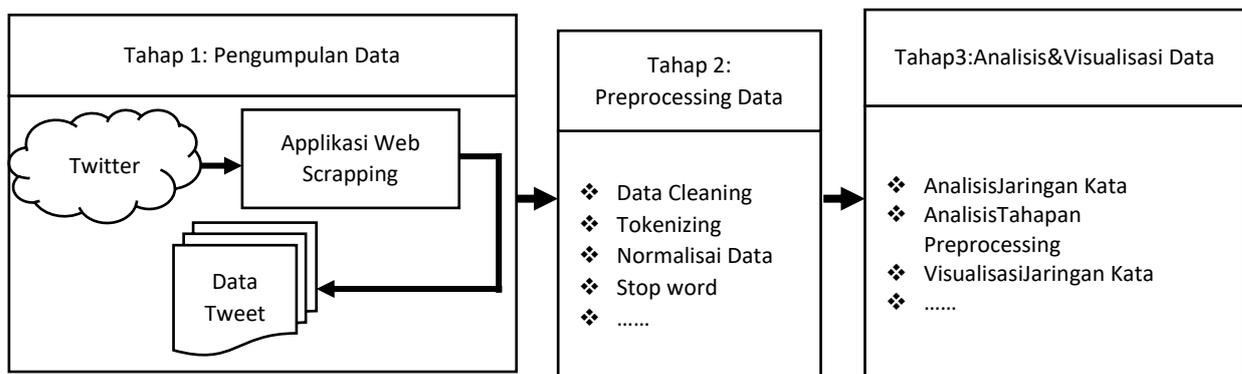
pandemi [7]. Brand Image dari suatu perguruan tinggi juga memiliki hubungan erat dengan media sosial [8]. Hal ini yang mendasari perguruan tinggi perlu memperhatikan fokus pembahasan dari stakeholder yang ada pada media sosial.

Media sosial YouTube, Facebook, Instagram, dan Twitter adalah media paling sering digunakan oleh masyarakat Indonesia [9], [10]. Twitter adalah media sosial yang dijadikan sebagai sumber data. Facebook tidak digunakan dikarenakan sifat jaringan sosial berbasis pertemanan yang membuat tidak semua informasi bisa diakses secara publik. Instagram dan youtube tidak digunakan karena fokus pada media sosial ini adalah foto dan video sedangkan data yang perlu diolah adalah berbentuk teks. Twitter menjadi salah satu media sosial yang aktif digunakan oleh masyarakat Indonesia dan berada pada posisi lima teratas sebagai platform media sosial yang paling banyak digunakan [11]. Indonesia memiliki sekitar 15,7 Juta Pengguna twitter pada juli 2021 [12]. Twitter merupakan media sosial untuk memposting opini dengan batasan 140 karakter dan pada september 2020 meningkat menjadi 280 karakter [13], [14]. Batasan karakter tersebut bukan merupakan kekurangan, karena dengan adanya batasan tersebut pengguna akan menyampaikan opininya secara singkat dan jelas dibandingkan jika tidak ada batasan karakter. Penarikan data tweet secara sederhana dapat dilakukan berdasarkan dari user atau keyword yang sudah ditentukan [14], agar dapat mengumpulkan banyak opini dari berbagai pengguna twitter maka lebih baik diambil berdasarkan keyword yang sudah ditentukan. Opini yang diambil pada penelitian ini adalah opini pengguna twitter tentang pendidikan tinggi. Penelitian ini hanya berfokus pada dua jenis saja yaitu pendidikan akademik dan pendidikan vokasi.

Ada dua tujuan dilakukan penelitian ini, tujuan pertama adalah untuk mengidentifikasi fokus pembahasan para pengguna twitter pada bidang pendidikan tinggi pada masa pandemi. Identifikasi hasil pembahasan ini dapat membantu pihak perguruan tinggi untuk memformulasikan strategi yang tepat untuk para stakeholder pada masa pandemi. Tujuan kedua adalah membandingkan struktur jaringan semantik dengan stemming dan tanpa stemming. Perbandingan ini dilakukan untuk mengetahui terdapat adanya pengaruh signifikan atau tidak pada hasil analisis jaringan. Hal ini dikarenakan terdapat penelitian yang menjelaskan bahwa hubungan semantik antara 2 kata akan memiliki dampak signifikan dalam pengolahan data teks [15]. Data tweet yang diekstrak bersumber dari tweet berbahasa Indonesia yang diambil pada awal bulan maret 2020 sampai pada akhir maret 2021. Ada dua macam data tweet yang diambil yaitu terkait dengan pendidikan akademik dengan kata kunci “universitas” dan jenis pendidikan vokasi dengan kata kunci “politeknik”.

2. METODOLOGI PENELITIAN

Subjek penelitian adalah seluruh masyarakat Indonesia yang menggunakan media sosial. Populasi penelitian adalah pengguna twitter di Indonesia. Ada tiga tahapan utama pada penelitian ini, yaitu pengumpulan data, preprocessing data serta analisis & visualisasi data. Rancangan alur penelitian yang dilakukan diperjelas pada **Gambar 1**.



Gambar 1. Diagram Alur penelitian visualisasi Data Tweet

2.1 Pengumpulan Data

Pengumpulan data yang dilakukan disini menggunakan pencarian berbasis keyword yang sudah ditentukan. Data diambil dari twitter dengan menggunakan bahasa Python yang memanfaatkan library Snsrape, library ini memanfaatkan bagian dari application programming interface (API) di Twitter untuk proses pengambilan data tweet. Library Snsrape membantu dalam mencari data tweet pada blok/rentang waktu yang sudah didefinisikan [16]. Data pada penelitian ini merupakan data selama 1 tahun saat masa pandemi yaitu 01 Maret 2020 sampai pada 31 Maret 2021. Data yang dikumpulkan ini memiliki kata kunci yang mewakili pendidikan akademik yaitu “universitas” dan untuk vokasi adalah “politeknik”. Kedua kata kunci ini diambil karena kalimat tersebut adalah yang paling mewakili bidang akademik dan vokasi yang sering menjadi bahasan pada media sosial.

Tweet yang diambil adalah tweet berbahasa Indonesia. Hal ini dilakukan dengan mendefinisikan bahasa id (singkatan dari bahasa Indonesia) pada query yang dijalankan dengan library Snsrape. Data tweet ini kemudian disimpan dalam format CSV untuk dapat diolah pada tahapan berikutnya. Jumlah total data yang didapatkan dan digunakan sebagai sumber penelitian yang dilakukan ini ditunjukkan pada **Tabel 1**.

Tabel 1. Hasil Pengumpulan Data Tweet

No.	Data Tweet	Jumlah Tweet
1.	Pendidikan Akademik (<i>Keyword: Universitas</i>)	328.835
2.	Pendidikan Vokasi (<i>Keyword: Politeknik</i>)	33.228
	Total	362.063

2.2 Preprocessing Data

Tahapan preprocessing data ini merupakan tahapan untuk menghilangkan noise. Pada data tweet noise yang dimaksud saat mengolah data teks adalah seperti emoticon, hashtags, simbol, retweet, singkatan dan berbagai kata tidak baku yang sulit dipahami [17]. Tahapan preprocessing secara umum terdiri dari case folding, tokenizing, filtering, and stemming [18]. Metode pada penelitian ini sesuai dengan urutannya adalah sebagai berikut: case folding untuk membuat semua huruf menjadi lowercase, remove special character (seperti: tab, new line, dan lainnya), remove non ASCII character (seperti: emoticon, karakter chinese, dan lainnya) remove twitter features yang tidak dibutuhkan (seperti: hashtags mention, link), remove incomplete link, remove karakter noise (punctuation, multiple space, karakter huruf tunggal), tokenizing, filtering (stopword menggunakan library nltk), normalisasi teks (menggunakan kamus slang dari github [19]), dan terakhir dilakukan stemming (menggunakan library Sastrawi). Tahapan preprocessing yang detail ini dilakukan karena ada penelitian menunjukkan bahasa informal (slang, penulisan yang tidak tepat, bahasa gaul) dan berbagai anomali pada data tweet akan mempengaruhi akurasi [15].

Hasil yang disimpan untuk digunakan pada tahapan berikutnya ada dua macam yaitu hasil teks normalisasi dan hasil stemming. Hal ini dibutuhkan karena penelitian ini ingin membandingkan hasil jaringan setelah dinormalisasi dengan yang sudah dilakukan stemming. Tahapan ini menghasilkan 4 dataset yang akan dianalisis dan divisualisasikan yaitu data hasil normalisasi teks untuk kata kunci “universitas”, hasil stemming untuk kata kunci “universitas”, hasil normalisasi teks untuk kata kunci “politeknik” dan hasil stemming untuk kata kunci “politeknik”. Sebelum analisis dan visualisasi data dilakukan, setiap data yang ada dilakukan kembali stopwords untuk memastikan bahwa kata yang ada pada setiap tweet sudah merupakan kata yang penting saja.

2.3 Analisis & Visualisasi Data

Tahapan terakhir ini adalah untuk melakukan analisis data serta membuat visualisasi struktur hubungan dari antar kata pada setiap tweet. Analisis ini dilakukan dengan melakukan proses feature extraction yang merubah data teks menjadi data numerik untuk dapat dipahami oleh komputer. Proses feature extraction ini dilakukan dengan menggunakan perhitungan frekuensi dari setiap kata. Proses selanjutnya adalah menggabungkan untuk tiap dua token dari setiap tweet yang saling berhubungan (disebut frasa bigram) [18]. Hal dilakukan agar dapat dianalisis hubungan semantik dari antar kata dan membuat visualisasi jaringan. Proses perhitungan untuk frekuensi setiap kata dan frekuensi frasa bigram untuk setiap dataset dilakukan dengan menggunakan **Persamaan 1**.

$$w_i = \sum term_i \quad (1)$$

Proses berikutnya adalah perhitungan untuk proporsi jumlah setiap tokennya. Hal ini dilakukan dengan melakukan pembagian dari hasil jumlah setiap token dibagi dengan jumlah seluruh data yang ada. Formula untuk menghitung proporsi kemunculan kata ini ditunjukkan pada **Persamaan 2**.

$$p_i = \frac{w_i}{\sum T_i} \quad (2)$$

Dimana:

Pi = Proporsi dari kemunculan token ke i.

wi = Jumlah dari token ke i.

Ti = Jumlah total seluruh token dari dataset.

Metode yang digunakan untuk melakukan evaluasi ukuran data adalah Shannon information theory. Shannon information theory adalah sebuah metode untuk mengukur secara tepat besaran informasi yang berguna [20]–[22]. Metode Shannon information theory ini digunakan untuk mengevaluasi ukuran informasi yang berguna dari setiap dataset. Formula untuk menghitung Shannon information theory ini ditunjukkan pada **Persamaan 3**.

$$H(p) = - \sum_i p_i \log_2(p_i) \quad (3)$$

Dimana:

H(p) = Entropy

Pi = Proporsi dari kemunculan token ke i.

Tahapan terakhir dari data frasa bigram yang sudah diolah maka dikembangkan jaringan sosial menggunakan Gephi untuk dapat dianalisis hubungan antar katanya. Gephi adalah aplikasi open-source untuk melakukan analisis grafik dan jaringan yang dikembangkan oleh SciencesPo dan Linkfluence di Prancis [23]. Gephi digunakan dalam penelitian ini untuk melakukan analisis dan visualisasi fitur struktur jaringan.

3. HASIL DAN PEMBAHASAN

Tahapan yang dilakukan setelah data tweet yang untuk dua kata kunci (universitas dan politeknik) berhasil dikumpulkan adalah preprocessing data. Tahapan preprocessing ini juga melakukan perhitungan jumlah seluruh kata dari data tiap tweet serta mencari tahu kata unik yang ada pada data tweet. Hasil preprocessing data yang memiliki jumlah kemunculan kata dibawah tiga dan juga frasa bigram dengan frekuensi dibawah tiga akan dihilangkan agar data yang diolah hanya data yang penting saja. Proses selanjutnya adalah perhitungan entropy (Shannon information theory) menggunakan Persamaan 3. Hasil preprocessing dan perhitungan entropi ditunjukkan pada **Tabel 2**.

Tabel 2. Hasil Preprocessing Dan Perhitungan Entropy

No.	Fokus	Jumlah Tweet	Jumlah Seluruh Kata	Kata Unik	Rata-Rata Jumlah	Entropy
1.	Universitas tanpa Stemming	328.835	4.337.752	53.593	80,94	11,70
2.	Universitas dengan Stemming		4.254.431	43.527	97,74	11,18
3.	Politeknik dengan Stemming	33.228	392.682	11.590	33,88	10,81
4.	Politeknik Tanpa Stemming		388.818	9.269	41,99	10,39

Hasil pada **Tabel 2** ini menunjukkan perbedaan yang signifikan antara hasil dengan menggunakan stemming dan tanpa menggunakan stemming. Hal ini dikarenakan fungsi dari stemming adalah mengkonversi kata berimbuhan menjadi kata dasar. Hasil perhitungan entropy juga menunjukkan bahwa dengan stemming maka besaran informasi yang penting jauh lebih kecil (perbedaan 0,52 pada kata kunci universitas dan perbedaan 0,42 pada kata kunci politeknik). Apabila hanya dilihat dari sisi jumlah seluruh kata, peningkatan jumlah rata-rata kata, besaran informasi maka jelas hasil preprocessing terlihat lebih baik, tetapi hal ini akan bergantung pada tujuan dari penelitian.

Penelitian ini melakukan visualisasi data opini dari para pengguna media sosial twitter, dimana tujuannya adalah melihat hubungan semantik dari setiap kata. Tujuan penelitian untuk melihat hubungan semantik ini berbeda dengan tujuan pembuatan model untuk classification atau clustering. Dimana fokus pembuatan model biasanya ada pada waktu proses dan akurasi dari model yang dibuat. Apabila dihubungkan dengan semantik maka hasil stemming tidak selalu dapat digunakan, ini dikarenakan stemming merubah kata menjadi kata dasar yang dapat membuat makna kata menjadi berbeda. Hasil stemming pada penelitian ini memang dapat memperkecil besaran informasi penting tetapi jika dilihat dari sisi semantik maka hal ini kurang tepat untuk dilakukan. Pendapat ini diperkuat dari hasil 10 data frekuensi tertinggi dari setiap kata dan frasa bigram untuk tiap dataset yang ditunjukkan pada **Tabel 3-4**.

Tabel 3. Sepuluh Kata Dengan Frekuensi Tertinggi Pada Tiap Dataset

No.	Universitas tanpa stemming		Universitas dengan stemming		Politeknik tanpa stemming		Politeknik dengan stemming	
	Kata	Freq	Kata	Freq	Kata	Freq	Kata	Freq
1.	universitas	346344	universitas	346768	politeknik	35005	politeknik	35076
2.	indonesia	46482	mahasiswa	49028	negeri	7160	negeri	7169
3.	mahasiswa	46476	indonesia	47245	mahasiswa	3613	daftar	4279
4.	fakultas	32506	fakultas	32670	masuk	2448	mahasiswa	3746
5.	negeri	30132	negeri	30225	daftar	2161	masuk	2870
6.	kuliah	23077	kuliah	26486	ptn	1787	ajar	2362
7.	jurusan	19266	terima	21246	kuliah	1778	terima	2248
8.	semoga	18607	jurusan	19914	universitas	1656	kuliah	1999
9.	pendidikan	18361	kampus	19791	jurusan	1648	jurusan	1821
10.	kampus	18298	didik	19789	pendaftaran	1637	ptn	1787

Tabel 3 menunjukkan adanya sedikit perubahan urutan dikarenakan jumlah frekuensi setelah dilakukan stemming yang mengubah menjadi kata dasar. Urutan yang berubah jika dilihat pada kata kunci 'universitas' adalah kata "mahasiswa", "kampus" yang meningkat, kemudian ada dua kata baru yang muncul yaitu "terima" dan "didik". Kata "terima" setelah di cek dari dataset yang tidak dilakukan stemming, kata ini berasal dari kata seperti "keterima", "diterima", "terima kasih" yang menunjukkan bahwa ada perubahan makna dari kata tersebut. Kata "didik" bisa berasal dari "kependidikan", "pendidik", "mendidik", "berpendidikan" yang jika diubah menjadi kata dasar, maka tidak dapat dipahami mana yang merupakan kata kerja atau kata benda. Hasil tabel tersebut juga menunjukkan metode stemming sastrawi ini merubah kata "jurusan" menjadi "jurus", yang artinya makna kata ini sudah sangat jauh berbeda.

Hal yang sama juga dapat terlihat pada kata kunci 'politeknik', dimana pada 10 data teratas yang perlu menjadi perhatian adalah kata "ajar", "daftar" dan "jurusan". Kata "ajar" bisa berasal dari kata seperti "pengajaran", "belajar" dimana merupakan dua hal yang berbeda (pengajaran menunjukkan kata benda, dan belajar merupakan kata kerja).

hasil pembahasan seperti yang ada pada **Tabel 3**, tetapi perlu diketahui jika kata tidak tepat maka akan berdampak pada frasa bigram karena merupakan penggabungan dua token yang berhubungan.

Tabel 4. Sepuluh Frasa *Bigram* Dengan Frekuensi Tertinggi Pada Tiap *Dataset*

No	Universitas Tanpa Stemming		Universitas dengan Stemming		Politeknik Tanpa Stemming		Politeknik dengan Stemming	
	Kata	Freq	Kata	Freq	Kata	Freq	Kata	Freq
1.	universitas indonesia	31156	universitas indonesia	31776	politeknik negeri	6797	politeknik negeri	6818
2.	universitas negeri	31156	universitas negeri	26467	mahasiswa politeknik	2017	mahasiswa politeknik	2107
3.	mahasiswa universitas	23710	mahasiswa universitas	24714	masuk politeknik	1420	masuk politeknik	1611
4.	fakultas universitas	20952	fakultas universitas	21173	politeknik jakarta	1005	daftar politeknik	1536
5.	masuk universitas	10834	kuliah universitas	11710	daftar politeknik	991	politeknik jakarta	1005
6.	universitas muhammadiyah	10702	masuk universitas	11231	ptn politeknik	899	kuliah politeknik	934
7.	kuliah universitas	10182	universitas muhammadiyah	10719	mahasiswa negeri	897	mahasiswa negeri	931
8.	universitas islam	9986	universitas islam	10044	politeknik bandung	889	ptn politeknik	917
9.	universitas brawijaya	9585	universitas brawijaya	9593	politeknik jember	884	politeknik bandung	894
10.	ilmu universitas	9022	ilmu universitas	9258	negeri politeknik	840	politeknik jember	885

Hasil analisis ini menunjukkan bahwa metode stemming dapat membuat makna semantik dari kata akan berubah, permasalahan ini bukan hanya dikarenakan penggunaan metode stemming yang kurang tepat (penelitian ini menggunakan sastrawi) tetapi memang tujuan dari metode stemming ini adalah untuk mereduksi dimensi dari data. Tujuan reduksi ini agar dapat mempercepat proses dan meningkatkan akurasi. Visualisasi hubungan antar kata dengan frasa bigram pada penelitian ini dibuat berdasarkan hasil dari masing-masing dataset. Setiap kata akan dijadikan node, setiap hubungan frasa bigram akan membuat sebuah hubungan kata yang disebut dengan edge dan setiap frekuensi kata dijadikan degree.

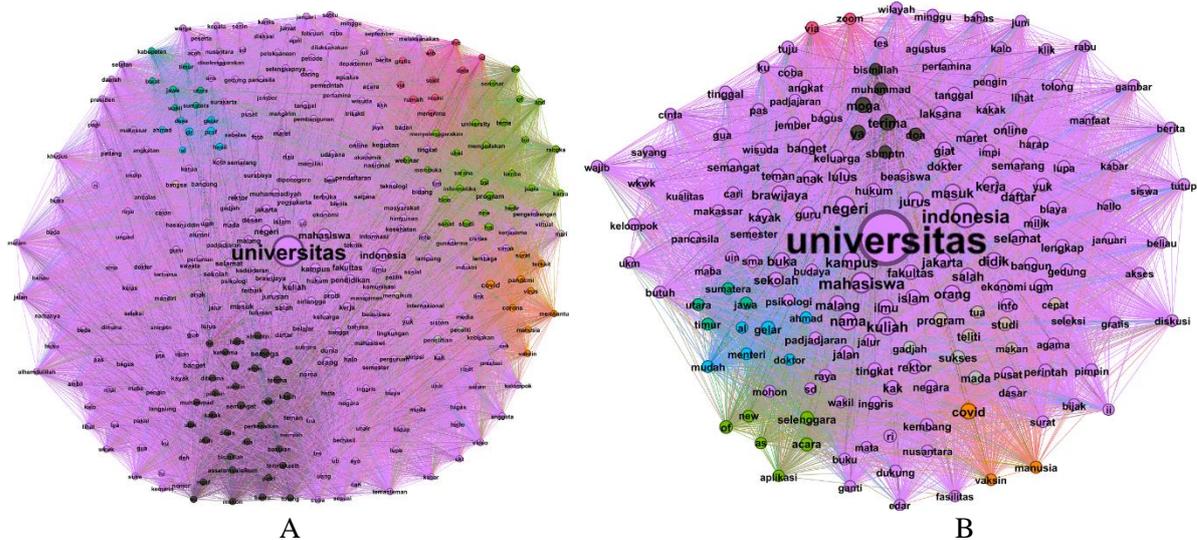
Frasa dari bigram yang dihasilkan dari preprocessing adalah data undirected (tidak berarah). Undirect yang dimaksud disini maka setiap frasa bigram ini bisa berasal dari dua arah seperti A-B dan B-A [23]. Visualisasi data yang ditampilkan pada penelitian ini adalah data dengan nilai degree minimal sebesar 2% dari nilai degree terbesar dari tiap dataset, ini dilakukan untuk memperjelas fokus pembahasan yang ada pada tiap dataset. Hasil pengolahan jaringan kata untuk tiap dataset dapat dilihat pada **Tabel 5**.

Tabel 5. Analisis Jaringan Kata

No.	Index	Universitas Tanpa Stemming	Universitas dengan Stemming	Politeknik Tanpa Stemming	Politeknik dengan Stemming
1.	Nodes	339	173	144	224
2.	Edges	44.315	12.425	4.941	9.689
3.	Density	0,774	0,835	0,48	0,388
4.	Avg. degree	261,445	143,642	68,625	86,509
5.	Modularity	0,106	0,032	0,112	0,068
6.	No. of communities	6	7	6	7
7.	Avg. path length	1,226	1,165	1,52	1,612
8.	Avg. clustering coefficient	0,826	0,88	0,65	0,611

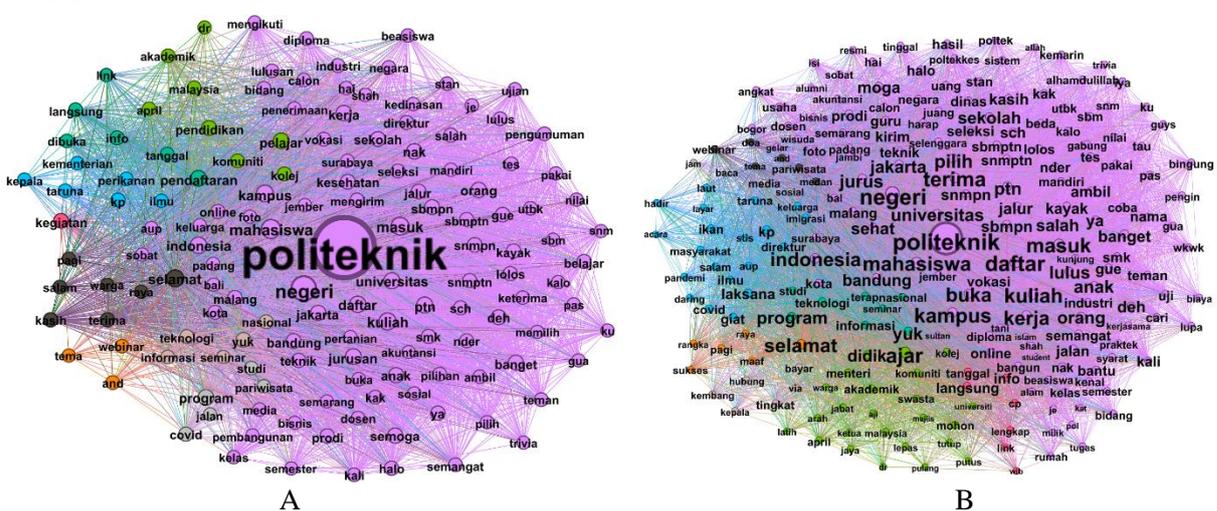
Nilai modularity menjelaskan tentang optimalisasi jumlah komunitas dalam sebuah jaringan [24]. Nilai modularity yang tinggi menunjukkan bahwa adanya koneksi padat antara node dalam kelompoknya sendiri tetapi memiliki koneksi yang jarang dengan kelompok yang berbeda. Tabel 5 menunjukkan bahwa nilai dari modularitas dari hasil stemming lebih kecil dibandingkan tanpa stemming. Artinya hasil preprocessing dengan stemming menunjukkan kurangnya keterhubungan didalam kelompok masing-masing tetapi terdapat hubungan yang tinggi dengan kelompok lain didalam jaringan. Density menunjukkan kepadatan jaringan, dimana semakin besar menunjukkan kemungkinan semua node yang terhubung dengan edge dan jumlah edge semakin mendekati/sama. Penelitian ini memiliki density diantara 0,8 untuk kata kunci universitas dan diantara 0,4 untuk kata kunci politeknik.

Dalam memvisualisasikan hubungan frasa bigram disini digunakan algoritma ForceAtlas2. ForceAtlas2 adalah algoritma untuk mengarahkan secara paksa node untuk memperlihatkan hubungan antara setiap node. Hasil visualisasi untuk hubungan frasa bigram ini ditunjukkan pada **Gambar 2** dan **Gambar 3**.



Gambar 2. Hasil Visualisasi Hubungan Semantik pada Keyword Universitas (A tanpa *Stemming* dan B dengan *Stemming*)

Visualisasi yang ditunjukkan pada **Gambar 2** memperlihatkan bahwa para pengguna twitter pada pembahasan terkait dengan universitas lebih mengarah kepada kehidupan di kampus seperti perkuliahan, tri dharma (penelitian, pengabdian dan pengajaran), jurusan/program studi terkait. Pada gambar dapat terlihat ada teridentifikasi beberapa komunitas yang mirip antara hasil visualisasi dengan stemming dan tanpa stemming. Komunitas pertama yang paling besar yang terkait dengan kehidupan kampus (“fakultas”, “mahasiswa”, “belajar”, “bangga”) yang diwarnai dengan warna ungu. Komunitas kedua terkait dengan doa untuk penerimaan ataupun salam (“semoga”, “diterima”, “lolos”, “assalamualaikum”). Komunitas ketiga yaitu pandemi (“covid”, “pandemi”, “corona”, “vaksin”) yang diwarnai dengan warna oranye. Komunitas yang tersisa adalah komunitas kecil yang berkaitan dengan penyelenggaraan/kalimat asing, gelar, lokasi dan pelaksanaan.



Gambar 3. Hasil Visualisasi Hubungan Semantik pada keyword Politeknik (A tanpa *Stemming* dan B dengan *Stemming*)

Visualisasi yang ditunjukkan pada **Gambar 3** menunjukkan pengguna twitter yang berkaitan dengan politeknik lebih fokus kepada penerimaan mahasiswa baru seperti jenis ujian, pendaftaran dan politeknik yang diminati. Dapat teridentifikasi beberapa komunitas yang mirip antara hasil visualisasi dengan stemming dan tanpa stemming. Komunitas yang paling besar berfokus pada penerimaan mahasiswa baru (“masuk”, “daftar”, “snmpn”, “sbmpn”, “jalur”) ditunjukkan warna ungu. Komunitas terbesar kedua adalah komunitas yang lebih ke arah pendidikan (“pendidikan”, “akademik”, “dr”) yang ditunjukkan warna hijau. Komunitas ketiga adalah komunitas institusi (“kementerian”, “taruna”, “kepala”) yang ditunjukkan warna biru, tetapi terdapat keanehan pada pemrosesan pengelompokan pada bagian stemming karena ada perpindahan dari kata yang terkait dengan pandemi (“daring”, “covid”, “pandemi”) ke komunitas berwarna biru. Hal ini dapat dipengaruhi oleh hasil frasa bigram yang berubah hubungan semantik antara katanya.

Penemuan menarik lainnya berdasarkan data dari kata kunci universitas ataupun politeknik hasil yang berkaitan dengan pandemi (“covid”, “daring”, “pandemi”, “vaksin”) tidak terlalu terlihat signifikan. Opini dari pengguna twitter tidak terlalu banyak yang menyampaikan pandemi, hal ini bisa dikarenakan sudah jelasnya dampak dari pandemi covid-19. Seperti terdapatnya perubahan kehidupan kampus ataupun seleksi mahasiswa baru yang dilakukan selama masa pandemi.

4. KESIMPULAN

Penelitian ini menunjukkan bahwa dengan menggunakan SNA dan membangun hubungan antara kata, kita dapat melihat berbagai macam topik/fokus yang sedang dibahas oleh pengguna media sosial twitter. Pemahaman topik fokus diskusi yang dibahas pada media sosial ini dapat membantu para stakeholder dan manajemen dari Pendidikan akademik ataupun vokasi dapat menentukan strategi bisnis untuk meningkatkan kepuasan civitas akademika ataupun memenuhi pencapaian visi, misi, tujuan dari perguruan tinggi. Hasil visualisasi data menunjukkan bahwa untuk pendidikan akademik khususnya pada universitas yang dijadikan kata kunci, fokus utama yang dibahas oleh para pengguna twitter adalah terkait dengan kehidupan kampus (perkuliahan, tri dharma, jurusan/prodi dan lainnya). Hal ini bisa dikarenakan para pengguna media sosial twitter yang sudah tidak menyukai kehidupan kampus secara online/daring. Strategi yang perlu dipertimbangkan jika pendidikan akademik masih memerlukan pelaksanaan secara daring maka perlu ada strategi yang membuat para mahasiswa merasakan kehidupan kampus walau secara daring. Pada visualisasi data politeknik yang menjadi fokus utama adalah terkait dengan proses penerimaan mahasiswa baru (pendaftaran, jenis ujian dan lainnya). Proses seleksi ini perlu menjadi perhatian dikarenakan apabila kampus vokasi menyiapkan strategi bisnis yang berkaitan dengan penerimaan mahasiswa baru maka akan terdapat peningkatan dalam jumlah mahasiswa. Penelitian ini juga membuktikan bahwa dengan menggunakan metode tambahan yang merubah makna kata seperti stemming akan membuat hasil pemaknaan semantik kata menjadi tidak tepat walaupun dapat menurunkan dimensi data dengan cukup signifikan. Penelitian ini hanya berfokus untuk melakukan analisis dan visualisasi hubungan antar kata dari twitter pada masa pandemi yaitu dari 01 maret 2020 sampai 31 maret 2021. Hasil ini perlu dibandingkan hasil data sebelum pandemi untuk mengetahui perubahan fokus tweet dari pengguna pada masa pandemi dan sebelum masa pandemi. Penelitian lebih lanjut pada media sosial lain seperti instagram, youtube dan facebook juga perlu dilakukan sebagai bentuk validasi hasil penelitian ini dengan hasil dari media sosial lain.

REFERENCES

- [1] A. G. Ekp, F. E. B. Unsyiah, K. Darussalam, B. Aceh, P. Covid, and D. I. Indonesia, “Merespon Nalar Kebijakan Negara Dalam Menangani Pandemi Covid 19 Di Indonesia,” *J. Ekon. dan Kebijak. Publik Indones.*, vol. 7, no. 1, pp. 36–53, 2020, doi: 10.24815/ekapi.v7i1.17370.
- [2] L. D. Herliandry, N. Nurhasanah, M. E. Suban, and H. Kuswanto, “Pembelajaran Pada Masa Pandemi Covid-19,” *JTP - J. Teknol. Pendidik.*, vol. 22, no. 1, pp. 65–70, 2020, doi: 10.21009/jtp.v22i1.15286.
- [3] N. N. S. Witari and J. Suryana, “Persepsi Mahasiswa Vokasi Terhadap Kegiatan Pembelajaran Mata Kuliah Teori Psikologi Komunikasi Pada Era Pandemi Covid 19,” *J. IKA*, vol. 18, no. 2, pp. 199–214, 2020.
- [4] D. C. U. Licharyani, P. G. A. P. Putra, R. V. H. Ginardi, and R. A. Sukmono, “Audit conformity for higher education using good university governance (Gug) principle,” *Proc. Int. Conf. Ind. Eng. Oper. Manag.*, vol. 0, no. March, pp. 2081–2089, 2020.
- [5] D. C. U. Licharyani, R. V. Hari Ginardi, R. Ambarwati, and M. T. Multazam, “Assessment for good university governance in higher education focus on align strategy business with it at big data era,” *J. Phys. Conf. Ser.*, vol. 1175, no. 1, 2019, doi: 10.1088/1742-6596/1175/1/012204.
- [6] S. Amin, “Academic Service Quality Improvement Strategies in Higher Education,” *J. Madaniyah*, vol. 7 Nomor 2, pp. 222–236, 2017.
- [7] M. Lucky and N. Rosmadi, “Penerapan Strategi Bisnis di Masa Pandemi Covid-19,” *IKRA-ITH Ekon.*, vol. 4, no. 1, pp. 122–127, 2021.
- [8] R. Anizir; Wahyuni, “Pengaruh Social Media Marketing Terhadap Brand Image Perguruan Tinggi Swasta Di Kota Serang,” *J. Sains Manaj.*, vol. 3, no. 2, p. 0, 2017.
- [9] Kastolani, “Understanding The Delivery of Islamophobic Hate Speech Via Social Media In Indonesia,” *Indones. J. Islam Muslim Soc.*, vol. 10, no. 2, pp. 247–270, 2020, doi: 10.18326/IJIMS.V10I2.247-270.
- [10] F. Anwar, “Perubahan dan Permasalahan Media Sosial,” *J. Muara Ilmu Sos. Humaniora, dan Seni*, vol. 1, no. 1, p. 137, 2017, doi: 10.24912/jmishumsen.v1i1.343.
- [11] A. BIN MUHAMMAD ALKATIRI, Z. NADIAH, and A. N. S. NASUTION, “Opini Publik Terhadap Penerapan New Normal Di Media Sosial Twitter,” *Cover. J. Strateg. Commun.*, vol. 11, no. 1, pp. 19–26, 2020, doi: 10.35814/coverage.v11i1.1728.
- [12] P. Patmawati and M. Yusuf, “Analisis Topik Modelling Terhadap Penggunaan Sosial Media Twitter oleh Pejabat Negara,” *Build. Informatics, Technol. Sci.*, vol. 3, no. 3, pp. 122–129, 2021, doi: 10.47065/bits.v3i3.1012.
- [13] F. R. Hartono, Y. A. Sari, and P. P. Adikara, “Pembangkitan Aturan Pengenalan Emosi Pada Twitter Menggunakan Metode Fuzzy-C Means,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 10, pp. 3258–3264, 2018.
- [14] K. Sussolaikah, “Pemanfaatan Packages Pada R Programming Untuk Crawling Data Pada Social Media,” *Build. Informatics, Technol. Sci.*, vol. 3, no. 3, pp. 203–206, 2021, doi: 10.47065/bits.v3i3.1035.
- [15] B. A. H. Murshed, H. D. E. Al-Ariki, and S. Mallappa, “Semantic analysis techniques using twitter datasets on big data: Comparative analysis study,” *Comput. Syst. Sci. Eng.*, vol. 35, no. 6, pp. 495–512, 2020, doi: 10.32604/CSSE.2020.35.495.



- [16] M. R. Adrian, M. P. Putra, M. H. Rafialdy, and N. A. Rakhmawati, “Perbandingan Metode Klasifikasi Random Forest dan SVM Pada Analisis Sentimen PSBB,” *J. Inform. Upgris*, vol. 7, no. 1, pp. 36–40, 2021, doi: 10.26877/jiu.v7i1.7099.
- [17] A. S. Aribowo and S. Khomsah, “Implementation Of Text Mining For Emotion Detection Using The Lexicon Method (Case Study: Tweets About Covid-19),” *Telematika*, vol. 18, no. 1, p. 49, 2021, doi: 10.31315/telematika.v18i1.4341.
- [18] M. A. Fauzi, D. C. Utomo, E. S. Pramukantoro, and B. D. Setiawan, “Automatic essay scoring system using N-GRAM and cosine similarity for gamification based elearning,” 2017, doi: 10.1145/3133264.3133303.
- [19] N. Aliyah Salsabila, Y. Ardhito Winatmoko, A. Akbar Septiandri, and A. Jamal, “Colloquial Indonesian Lexicon,” *Proc. 2018 Int. Conf. Asian Lang. Process. IALP 2018*, pp. 226–229, 2019, doi: 10.1109/IALP.2018.8629151.
- [20] C. J. Maley, “Continuous Neural Spikes and Information Theory,” *Rev. Philos. Psychol.*, vol. 11, no. 3, pp. 647–667, 2020, doi: 10.1007/s13164-018-0412-5.
- [21] J. A. Danowski, B. Yan, and K. Riopelle, *A semantic network approach to measuring sentiment*, vol. 55, no. 1. Springer Netherlands, 2021.
- [22] J. A. Danowski and K. Riopelle, “Scaling constructs with semantic networks,” *Qual. Quant.*, vol. 53, no. 5, pp. 2671–2683, 2019, doi: 10.1007/s11135-019-00879-5.
- [23] Z. Hou, F. Cui, Y. Meng, T. Lian, and C. Yu, “Opinion mining from online travel reviews: A comparative analysis of Chinese major OTAs using semantic association analysis,” *Tour. Manag.*, 2019, doi: 10.1016/j.tourman.2019.03.009.
- [24] M. Koponen, M. A. Asikainen, A. Viholainen, and P. E. Hirvonen, “Using network analysis methods to investigate how future teachers conceptualize the links between the domains of teacher knowledge,” *Teach. Teach. Educ.*, 2019, doi: 10.1016/j.tate.2018.12.010.