Building of Informatics, Technology and Science (BITS)
Volume 3, No 3, December 2021, Page 122−129
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v3i3.1012

Analisis Topik Modelling Terhadap Penggunaan Sosial Media Twitter oleh Pejabat Negara

Patmawati*, Muhammad Yusuf

Teknik Informatika, Universitas Amikom Yogyakarta, Yogyakarta, Indonesia
Email: [1,*]patmawati@students.amikom.ac.id, [2]yusuf.1175@students.amikom.ac.id
Email Penulis Korespondensi: patmawati@students.amikom.ac.id
Submitted: **05/12/2021**; Accepted: **25/12/2021**; Published**: 31/12/2021**

Abstrak−Social media mengalami perkembangan pesat hingga saat ini. Social media banyak memberikan kemudahan bagi manusia untuk dapat terhubung satu sama lain. Salah satu social media yang banyak digunakan adalah Twitter. Kemudahan penggunaan membuat social media ini banyak digunakan dikalangan masyarakat, termasuk kalangan para pejabat negara. Kalangan pejabat negara menggunakan twitter untuk meyampaikan kebijakan, pendapat serta melakukan interaksi dengan masyarakat. Dengan melakukan analisis topik terhadap tweets yang dibagikan oleh pejabat negara, kita dapat mengetahui topik relevan yang dibahas oleh pejabat negara tersebut. Kita dapat mengetahui fokus perhatian dari para pejabat negara melalui modelling topik. *Latent Dirichlet Allocation* (LDA) merupakan salah satu metode permodelan topik yang menemukan pola tertentu pada dokumen dan menghasilkan beberapa macam topik yang berbeda. Tweet dari akun @jokowi di kumpulkan menggunakan teknik scrapping. Hasil tweet collection tersebut kemudian dipreprocessing untuk selanjutnya dianalisis menggunakan metode *Latent Dirichlet Allocation* (LDA). Hasil dari model analisis di evaluasi menggunakan perhitungan perplexity serta coherence score. Evaluasi model menghasilkan nilai perplexity sebesar -8.069 dan coherence score sebesar 0.375 untuk jumlah topik sebanyak 7. Hal tersebut menunjukkan bahwa model yang digunakan baik untuk menganalisis dan mencari topik dalam tweets para pejabat negara.

Kata Kunci: Topik Modelling; Latent Dirichlet Allocation (LDA); Perplexity; Coherence Score

Abstract−Social media is experiencing rapid development until now. Social media makes it easy for humans to be able to connect with one another. One of the social media that is widely used is Twitter. The ease of use makes social media widely used among the public, including state officials. State officials use Twitter to convey policies, opinions and interact with the public. By conducting a topic analysis of tweets shared by state officials, we can find out the relevant topics discussed by state officials. We can find out the focus of attention of state officials through topic modeling. Latent Dirichlet Allocation (LDA) is a topic modeling method that finds certain patterns in documents and produces several different topics. Tweets from the @jokowi account are collected using a scrapping technique. The results of the tweet collection are then preprocessed for further analysis using the Latent Dirichlet Allocation (LDA) method. The results of the analysis model are evaluated using perplexity calculations and coherence scores. The evaluation of the model resulted in a perplexity value of -8.069 and a coherence score of 0.375 for a total of 7. This shows that the model used is good for analyzing and finding topics in tweets of state officials.

Keywords: Modelling Topic; Latent Dirichlet Allocation (LDA); Perplexity; Coherence Score

1. PENDAHULUAN

Perkembangan sosial media semakin pesat hingga saat ini. Hal tersebut dikarenakan social media dapat memberikan kesempatan kepada masyarakat untuk terhubung kedunia online dalam bentuk politik, personal ataupun kegiatan bisni. Berbagai macam sosial media dapat digunakan oleh masyarakat untuk berkomunikasi ataupun menyampaikan aspirasi dan opini. Salah satu media sosial yang banyak digunakan yaitu Twitter. Penggunaan twitter di Indonesia mengalami peningkatan dari tahun 2020. Seperti yang dilansir pada situs katadaata.co.id, pengguna twitter di Indonesia pada Juli 2021 sebanyak 15.7 juta.

Twitter memiliki kemampuan koneksi yang luas sehingga platform ini banyak dimanfaatkan oleh beberapa kalangan, diantaranya kalangan pejabat negara. Hampir banyak pejabat negara yang memiliki akun twitter. Tujuan umum dari adanya akun tersebut adalah sebagai media untuk mempromosikan kebijakan politik, pendapat politik serta dapat berinteraksi dengan masyarakat. Masyarakat dapat memberikan pendapat atau komentar yang dikeluarkan oleh pejabat negara melalui twitter. Dan sebaliknya, para pejabat negara pun dapat memberikan tanggapan dari pembacanya dengan cepat [1]. Biasanya tweet yang dibagikan oleh para pejabat negara adalah informasi terkait keadaan daerah yang dipimpin. Dapat juga berisikan upaya-upaya yang telah dilakukan oleh para pejabat negara selama menjabat. Edukasi, sosialisasi serta publikasi kegiatan pun dapat diinformasikan melalui twitter. Setiap tweets yang dibagikan oleh para pejabat negara, tentunya memiliki topik-topik tertentu. Untuk itu diperlukan suatu metode yang dapat membantu masyarakat dalam memahami topik dari setiap tweet yang dibagikan oleh para pejabat.

Permodelan Topik merupakan salah satu metode analisis topik terhadap tweet-tweet yang dibagikan ataupun dokumen lainnya. Tujuannya yaitu untuk memperoleh sebuah topik utama dari sebuah tweets. Dengan begitu, kita dapat mengetahui gambaran umum terhadap isi dari topik utama. Metode yang dapat digunakan untuk permodelan topik ada banyak, diantaranya yaitu *Latent Dirichlet Allocation* (LDA).

Metode LDA adalah sebuah metode text mining dalam menemukan suatu pola tertentu pada dokumen dengan menghasilkan beberapa macam topik yang berbeda, sehingga tidak secara spesifik mengelompokkan dokumen kedalam sebuah topik tertentu [2]. Metode latent dirichlet allocation (LDA) memiliki performa yang unggul bila dibandingkan dengan metode pemodelan topik yang lain serta dapat diimplementasikan untuk mengidentifikasi topik dalam jurnal ilmiah, klasifikasi, dan pengelompokan [3].

Building of Informatics, Technology and Science (BITS)
Volume 3, No 3, December 2021, Page 122−129
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v3i3.1012

Penelitian terkait topik modelling menggunakan metode LDA (*Latent Dircehlet Allocation*) telah banyak dilakukan oleh peneliti-peneliti sebelumnya. Berikut beberapa penelitian yang dapat dijadikan acuan dalam penelitian ini, diantarnya penelitian yang berjudul "*Latent Direchlet Allocation (LDA) Untuk mengetahui Topik Pembicaraan Warganet Tentang Omnibus Law*". Dalam penelitian tersebut ditemukan jumlah topik sebanyak 5 dengan nilai koherensi sebesar 0.5644 [4].

Penelitian lainya berjudul "*Permodelan topik Pengguna Twitter Mengenai Aplikasi Ruang Guru*". Penelitian ini melakukan proses klasifikasi terhadap pendapat para pengguna terkait layanan ruangguru. Hasil penelitian membentuk 28 topik dengan menggunakan metode *Latent Direchlet Allocation* [5].

Penelitian dengan judul *Clustering topik pada data sentimen BPJS Kesehatan menggunakan metode Latent Dirichlet Allocation*. Hasil dari penelitian ini diperoleh jumlah topik sebanyak 2 dengan nilai perplexity sebesar 6.0907, nilai alpha sebesar 0.01 dan nilai beta sebesar 0.1 untuk sentimen positif. Untuk sentimen negatif, nilai perplexity 6.7364, nilai alpha sebesar 0.001 dan nilai beta sebesar 0.1. Sedangkan untuk sentimen netral diperoleh nilai perplexity 6.2094, nilai alpha sebesar 0.001 dan nilai beta sebesar 1. Perplexity merupakan pengukuran untuk mengevaluasi kinerja model LDA atau ketepatan informasi dari proses permodelan topik pada dokumen [6].

Penelitian terkait lainnya dengan judul *A Text Mining Research Based on LDA Topic Modelling*. Penelitian ini menggunakan metode LDA untuk mencari topik modelling terhadap dua tipe dokumen, yaitu Wikipedia dan Tweet dari Twitter. Secara keseluruhan, penelitian ini menjelaskan secara umum tentang proses text mining menggunakan metode LDA, pre-processing data, memodelkan algoritma serta evaluasi model [7].

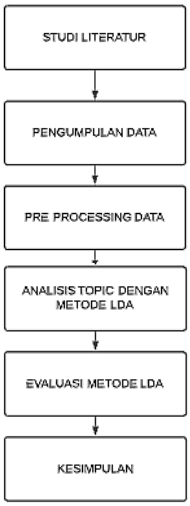
Selanjutnya peneltian berjudul *Analisis Topik Penelitian Kesehatan di Indonesia Menggunakan Metode Topic Modelling LDA (Latent Direchlet Allocation)*. Penelitian ini melakukan identifikasi topik terhadap judul - judul penelitian pada bidang kesehatan di Indoensia. Judul-judul tersebut diambil dari jurnal SINTA. Evaluasi model dilakukan dengan melihat nilai kohesi yang dihasilkan untuk setiap topik. Hasil pengujian menyimpulkan bahwa metode LDA sangat baik dilakukan untuk analisis topik terhadap penelitian – penelitian bidang kesehatan [8].

Berdasarkan pemaparan beberapa penelitian diatas, maka peneliti melakukan analisis topik terhadap tweet yang dibagikan oleh pejabat negara. Metode analisis topik yang dibangun adalah LDA (*Laten Directhlet Allocation)* dengan evaluasi metode menggunakan perhitungan nilai perplexity serta coherence untuk mencari keterkaitan dari uraian probabilitas kata-kata yang ditemukan dalam penyusunan suatu topik [9]. Hasil dari analisis topik modelling ini, diharapkan dapat memudahkan masyarakat dalam memahami isi tweet yang dibagikan oleh para pejabat. Sehingga masyarakat dapat mengetahui kinerja serta fokus perhatian dari para pejabat negara.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Tahapan penelitian yang digunakan pada penelitian ini dapat dilihat pada Gambar 2.1 Tahapan Metode penelitian, sebagai berikut:



Gambar 1. Tahapan Metode Penelitian

Tahapan penelitian dimulai dari studi literatur. Pada tahap ini, peneliti melakukan literatur review terhadap jurnal-jurnal atau pun pustaka-pustaka yang relevan dengan judul penelitian. Studi literatur dilakukan untuk mencari bahan acuan dalam pembuatan serta pendukung untuk melakukan penelitian ini. Selanjutnya dilakukan pengumpulan data untuk bahan proses topik modelling. Setelah data dikumpulkan, dilanjutkan degan tahap pre processing, yaitu pembersihan data. Data yang telah bersih kemudian di analisis menggunakan metode LDA. Berikutnya, tahapan evaluasi dimana model yang digunakan diuji untuk mengetahui keakurasian model. Dan terakhir tahapan pengambilan kesimpulan terkait hasil penelitian.

Building of Informatics, Technology and Science (BITS)
Volume 3, No 3, December 2021, Page 122−129
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v3i3.1012

2.2 Pengumpulan Data

Pengumpulan data dilakukan dengan teknik scrapping tweets pada salah satu akun twitter pejabat negara. Untuk akun pejabat negara yang digunakan adalah Presiden Republik Indonesia, Ir. H. Joko Widodo, dengan username twitter @jokowi. Teknik scrapping dilakukan dengan menggunakan library tweepy serta API twitter. Hasil scrapping tweets disimpan dalam bentuk file CSV (*Comma Separated Value*).

2.3 Preprocessing Data

Setelah melakukan pengumpulan data, langkah selanjutnya yaitu preprocessing data. Tujuannya yaitu menghilangkan atau memperbaiki beberapa character untuk keperluan proses analisis topik. Tahapan preprocessing penting dalam menyusun teks yang tidak terstruktur serta mejaga keyword yang dapat berguna untuk mewakili topik [10]. Tahapan preprocessing diantaranya yaitu:

a) *Casefolding*, mengubah seluruh characte pada tweets menjadi huruf kecil.
b) *Removal Punctuation*, menghapus character, URL, mention, hastag, angka, spasi ganda, serta tanda baca yang tidak digunakan dalam proses analisis topik.
c) *Tokenizing*, memanggal kalimat tweet menjadi bagian-bagian atau kata – kata yang lebih kecil
d) *Normalisasi*, memperbaiki kata yang disingkat.
e) *Stopword*, melakukan penghapus kata yang tidak mengandung makna arti lebih
f) *Stemming*, menghapus suffiks dalam suatu kata.

2.3 Membangun Model LDA

Setelah melewati preprocessing data, selanjutnya data dianalisis menggunakan model LDA. LDA merupakan salah satu metode machine learning termasuk unsupervised learning yang digunakan dalam pengelompokkan data kedalam kelas, meringkas serta melakukan proses data yang ukurannya besar [11]. Biasanya digunakan untuk untuk mengenali struktur dari topik laten pada sebuah dokument tekstual [7]. Ataupun juga dapat digunakan dalam untuk retrieval field information, permodelan dokumen dan juga klasifikasi.

LDA menganggap setiap dokumen dalam corpus adalah topik laten yang bercampur secara acak. Dan setiap topik laten merupakan character atau kata yang didistribusikan dan mewakili keseluruhan kata dalam dokumen. Topik laten dapat dihasilkan dari kumpulan kata pada dokumen dengan nilai proporsi yang berbeda [12]. Data yang ada terlebih dahulu dibuat menjadi corpus atau *Dictionary* dengan memanfaatkan modul library *Gensim* pada Python, yaitu *gensim.corpora.* Dictionary dibuat untuk melatih model LDA serta mempercepat proses pemprosesan data. Selanjutnya kata-kata dalam dictionary tersebut dipresentasikan dalam bentuk matriks atau disebut dengan istilah *bag of words. Bag of words* melakukan perhitungan jumlah kemunculan setiap kata pada dictionary. Nilai jumlah kemunculan tersebut yang dijadikan untuk pemprosesan topik modelling. *Bag of word* dapat mempresentasikan setiap dokumen melalui pengabaian urutan kata-kata dalam dokumen dan struktur sintaksis dari dokumen serta kalimat [13].

Kemudian membangun model LDA dengan menggunakan modul *gensim.model* dengan pemanggilan fungsi yaitu *LdaModel*. Hasil output dari model dianalisis dan kemudian divisualisasikan secara interaktif melalui library *PyLDAvis python. PyLDAvis* dirancang untuk membantu peneliti dalam menafsirkan topik dari model topik yang sesuai dengan kumpulan teks data [4]. Melalui pyLDAvis, persebaran kata serta relevansinya untuk tiap topik dapat terlihat. Hasil visualisasi permodelannya ditampilkan dalam dashboard berbasis web interaktif yang dapat mudah dipahami oleh penggunanya.

2.4 Evaluasi Model

Hasil dari model yang dibuat kemudian di evaluasi dengan melakukan perhitungan nilai perplexity serta coherence. Evaluasi dilakukan untuk mengetahui keakuratan dari validasi topik. Perplexity adalah ukuran kinerja dari permodelan bahasa berdasarkan probabilitas rata-rata yang telah dikembangkan. Dapat pula digunakan dalam membandingkan berbagai macam permodelan bahasa, hanya saja diperlukan teks yang serupa sebagai teks data testing. Perplexity juga merupakan sebuah metric yang digunakan untuk menguji ketepatan suatu informasi darai sebuah dokumen terhadap topik yang dihasilkan. Dan Perhitungannya dapat dilakukan dengan menentukan kemungkinan dari log teks suatu dokumen yang tidak terlihat [14]. Nilai perplexity yang lebih rendah memiliki arti bahwa kemampuan generalisasi yang lebih baik dalam artian, model yang digunakan semakin baik. Berikut adalah perhitungan perplexity [15]:

$$\text{per}(D_{test})=\exp\left\{\frac{-\sum_d \log p(W_d)}{\sum_d N_d}\right\} \quad (1)$$

Keterangan:

$p(W_d)$: peluang dari jumlah kata
N_d : Total jumlah kata dalam suatu dokumen ke-d

Cara lain untuk mencari jumlah topik terbaik yaitu dengan menggunakan coherence score. Coherence Score atau disebut dengan Topic Score merupakan bentuk evaluasi topik yang lebih mudah dalam interpretasi oleh manusia. Coherence Score digunakan untuk mengukur nilai suatu topik dengan melalui pengukuran tingkat kesamaan semantik dalam kata-kata yang terdapat pada topik [16]. Nilai topik coherence yang tinggi mengindikasikan bahwa model yang digunakan adalah baik [17].

Building of Informatics, Technology and Science (BITS)
Volume 3, No 3, December 2021, Page 122−129
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v3i3.1012

3. HASIL DAN PEMBAHASAN

3.1 Hasil Data Collection

Proses pegambalian data twitter melalui teknik scrapping dengan memanfaatkan API keyTwitter. Data tweet yang digunakan diambil dari akun twitter @jokowi pada periode 21 Juni 2021 hingga 25 Desember 2021. Data yang didapatkan terdiri dari data tanggal dan data tweet. Total tweet yang berhasil didapatkan sebanyak 500 tweets. Data tersebut kemudian dipreprocessing dengan beberapa tahapan preprocessing, hingga menghasilkan data bersih yang siap untuk di analisis. Hasilnya dapat dilihat pada Tabel 1 berikut:

Tabel 1. Data Tweets

Tanggal	Text
2021-12-25 01:02:17+00:00	pandemi bekap hidup saudara saudara umat kristiani natal batas harap kurang gembira ceria raya orang sayang
2021-12-24 06:20:37+00:00	resmi gedung kantor dewan masjid indonesia jakaa harap dmi semangat jadi ibadah pusat didik dakwah musyawarah bangun satu sea tingkat sejahtera masyarakat
2021-12-23 05:11:59+00:00	kerja gotong royong upaya putus rantai sebar virus covid jaga lonjak libur
2021-12-23 05:11:58+00:00	pek indonesia suntik juta dosis vaksin masyarakat target anak usia vaksinasi capai hasil gotong royong
2021-12-22 07:09:53+00:00	nu milik potensi rangka perata ekonomi umat kuat anak muda santri kualitas kompetensi rajut lokomotif tarik gerbong sejahtera
2021-12-22 07:09:51+00:00	buka muktamar nahdlatul ulama lampung siang terima kasih nu awal jalan perintah senantiasa depan bangsa toleransi majemuk pancasila uud sea nkri

Hal pertama yang dilakukan adalah mengubah data tweet diatas kedalam bentuk dictionary serta *bag of words*. Hasil pengubahan kedalam Dictionary terdapat 3056 kata-kata unique. Selanjutnya bag of word melakukan pengindeksan setiap kata serta menghitung kemunculan atau probabilitas berdasarkan unique kata yang ditentukan pada proses dictionary. Setelah itu dilanjutkan dengan permodelan Latent Direchlet Allocation (LDA).

3.2 Hasil Permodelan Topik

Parameter yang dijadikan acuan untuk menghasilkan model topik yang terbaik yaitu number of topics dan words in topic. Number of topic adalah jumlah topik yang diperoleh dalam satu dokumen, sedangkan words in topic merupakan jumlah kata yang menyusun topik. Dalam penelitian ini digunakan *Number of topic* = 7 dan *word of topic* = 10. Output dari permodelan topik menghasilkan nilai probabilitas dari beberapa kata. Probabilitas merupakan frekuensi kemunculan kata pada dokumen. Pemilihan kata diambil sebanyak jumlah kata (*word of topic*) yang memiliki probabilitas tertinggi. Berikut keluaran topik yang dihasilkan dari permodelan topik

Tabel 2. Output Permodelan Topik

Topik	Probabilitas * Kata
0	'0.015*"vaksinasi" + 0.013*"covid" + 0.012*"jawa" + 0.010*"masyarakat" + ' '0.010*"daerah" + 0.009*"timur" + 0.009*"pagi" + 0.008*"pemerintah" + ' '0.007*"kegiatan" + 0.007*"meninjau"
1	'0.028*"indonesia" + 0.015*"selamat" + 0.012*"tokyo" + 0.011*"medali" + ' '0.009*"negara" + 0.008*"atlet" + 0.007*"paralimpiad" + 0.005*"air" + ' '0.005*"meraih" + 0.005*"masyarakat"'
2	'0.009*"masyarakat" + 0.009*"ha" + 0.008*"kab" + 0.008*"jalan" + ' '0.007*"nasion" + 0.006*"dibangun" + 0.006*"bendungan" + 0.006*"indonesia" + ' '0.006*"pemerintah" + 0.006*"rp"'
3	'0.009*"covid" + 0.007*"pasar" + 0.006*"jakaa" + 0.006*"negeri" + ' '0.005*"pedagang" + 0.005*"kaki" + 0.005*"pandemi" + 0.005*"warung" + ' '0.004*"bantuan" + 0.004*"siang"
4	'0.019*"indonesia" + 0.018*"vaksinasi" + 0.017*"covid" + 0.015*"vaksin" + ' '0.010*"juta" + 0.010*"dosi" + 0.010*"kesehatan" + 0.010*"pemerintah" + ' '0.008*"pandemi" + 0.007*"masyarakat"
5	'0.010*"pandemi" + 0.006*"indonesia" + 0.006*"dunia" + 0.005*"rumah" + ' '0.005*"negara" + 0.005*"kesehatan" + 0.005*"ekonomi" + 0.005*"bendungan" + ' '0.004*"ktt" + 0.004*"covid"
6	'0.009*"pagi" + 0.007*"jakaa" + 0.007*"covid" + 0.006*"indonesia" + ' '0.005*"tanah" + 0.005*"rumah" + 0.005*"masyarakat" + 0.004*"meninjau" + ' '0.004*"meresmikan" + 0.004*"vaksinasi"

Building of Informatics, Technology and Science (BITS)
Volume 3, No 3, December 2021, Page 122−129
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v3i3.1012

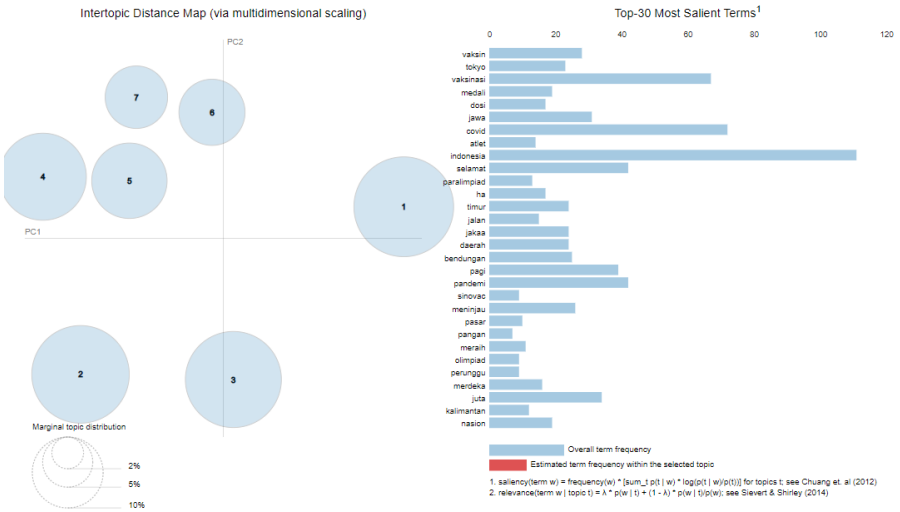
Dari hasil permodelan topik diatas, peneliti melakukan analisis untuk menyusun topik berdasarkan nilai probabilitas kata yang muncul dari model yang telah dibuat. Berikut adalah hasil analisis peneliti dari tabel hasil permodelan topik diatas.

Tabel 3. Hasil Analisis dari Output Permodelan Topic

Topic	Hasil Analisis
Topic 0	Peninjuan vaksinasi covid pada masyarakat di jawa
Topic 1	Ucapan selamat pada altet indonesia peraih medali pada olimpiade tokyo
Topic 2	Pembangunan jalan dan bendungan oleh pemerintah untuk masyarakat
Topic 3	Pedagang di pasar butuh bantuan saat pandemi
Topic 4	Penambahan dosis vaksin untuk vaksinasi demi kesehatan
Topic 5	Kesehatan dan ekonomi indonesia pada saat pandemi covid
Topic 6	Meninjau dan meresmikan program vaksinasi di masyrakat

3.3 Visualisasi Permodelan Topik

Output dari permodelan topik dapat divisualisasikan menggunakan library *pyLDAvis* dan hasilnya dapat dilihat pada gambar 2 berikut.



Gambar 2. Visualisasi Permodelan Topik

Pada gambar diatas, terdapat bentuk lingkaran yang menggambarkan topik. Semakin besar bentuk lingkarannya menandakan bahwa topik tersebut sangat berpengaruh atau sering muncul di dokumen. Sedangkan kata-kata yang ada di panel kanan merupakan kata – kata dominan yang dibahas dalam topik. Terdiri dari 30 terms atau kata dan menampilkan persentase dari kata tokens untuk setiap topik. Kata – kata tersebut terdiri dari *vaksin, tokyo, vaksinasi, medali, dosis, jawa, covid, atlet, indonesia, selamat, paralimpiad, ha, timur, jalan, jakaa, daerah, bendungan, pagi, pandemi, sinovac, meninjau, pasar, pangan, meraih, olimpiad, perunggu, merdeka, juta, kalimantan, nasion.* Apabila kata-kata pada panel disamping dipilih, maka bentuk lingkaran mengalimi perubahan bentuk. Semakin besar bentuk lingkaran, menunjukkan kata tersebut muncul dominan pada topik tersebut.

3.4 Evaluasi Permodelan Topik

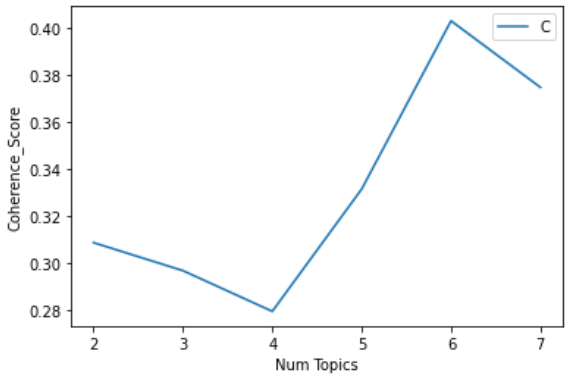
Tahapan ini dilakukan untuk memastikan bahwa model yang dibentuk untuk topik modelling dokumen menghasilkan nilai probabilitas yang tertinggi. Evaluasi dilakukan melalui dua metode yaitu dengan metode *Perplexity* dan dengan metode *Score Topic Cohorence.* Parameter yang mejadi acuan dalam evaluasi model yaitu jumlah topik. Berikut adalah tabel nilai perplexity dan coherence score yang diperoleh dalam penelitian ini:

Tabel 4. Nilai Evaluasi *Perplexity* dan Nilai *Coherence Score*

Jumlah Topik	Nilai *Perplexity*	*Coherence Score*
7	-8.06899991721538	0.3745496387523902
6	-8.025974717151541	0.40277582878256496
5	-7.976868943664449	0.33135125281623257
4	-7.913609516505872	0.2794537291050216
3	-7.840355512228399	0.29673333225230447
2	-7.772051970871911	0.30861439227062853

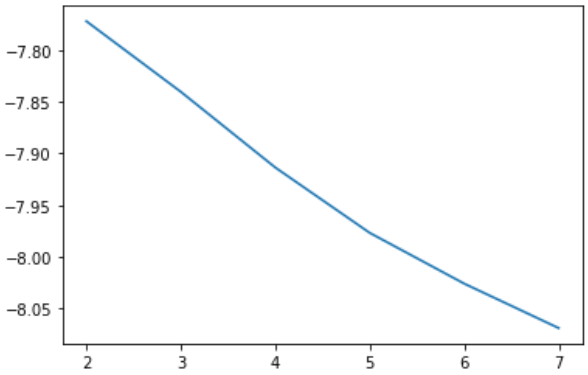
Building of Informatics, Technology and Science (BITS)
Volume 3, No 3, December 2021, Page 122−129
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v3i3.1012

Tabel diatas yang terdiri dari jumlah topik, nilai perplexity dan coherence score dapat divisiualisasikan dengan menggunakan grafik seperti yang dilihat pada grafik berikut ini:



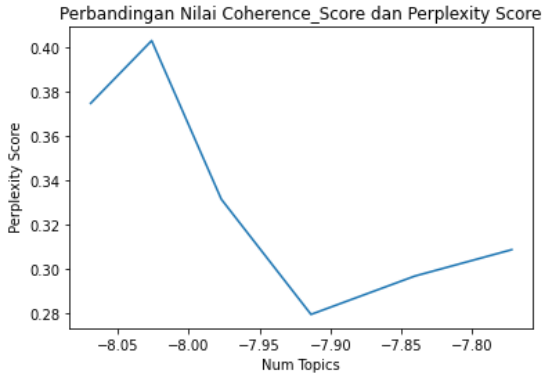
Gambar 3. Grafik Nilai Coherence Score

Dari gambar grafik diatas diperoleh nilai coherence score tertinggi berada pada jumlah topic sebanyak 6, dan paling rendah pada jumlah topik sebanyak 4. Dari grafik tersebut dapat ditarik kesimpulan bahwa, banyak jumlah topik yang digunakan dalam model berpengaruh terhadap nilai coherence score.



Gambar 4. Grafik Nilai Perplexity

Dari grafik diatas dapat disimpulkan semakin kecil jumlah topik, maka semakin besar nilai perplexity. Diperoleh nilai perplexity terbaik berada pada jumlah topik sebanyak 7, dikarenakan semakin kecil nilai perplexity, maka semakin baik performa dari model yang dibuat. Visualisasi perbandingan nilai perplexity terhadap coherence score dapat dilihat pada grafik berikut ini.



Gambar 5. Grafik Perbandingan Nilai Coherence Score dan Perplexity Score

Grafik diatas merupakan visualisasi keterkaitan antara nilai perplexity dan coherences score berdasarkan data tabel 4 diatas.

3.5 Pengelompokkan Tweets

Berikut adalah pengelompokkan beberapa dokumen atau tweets dari data yang telah dipreprocessing, dengan menggunakan model LDA yang dibentuk diatas. Dokumen atau tweets yang digunakan sebanyak 6 dokumen atau tweets. Dan hasilnya dapat dilihat pada tabel 5 berikut

Building of Informatics, Technology and Science (BITS)
Volume 3, No 3, December 2021, Page 122−129
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v3i3.1012

Tabel 5. Klasifikasi Dokument atau Tweet

Document/Tweet	Topik	Topic Cohorence Score
pandemi bekap hidup saudara saudara umat kristiani natal batas harap kurang gembira ceria raya orang sayang	4	0.949495
resmi gedung kantor dewan masjid indonesia jakaa harap dmi semangat jadi ibadah pusat didik dakwah musyawarah bangun satu sea tingkat sejahtera masyarakat	3	0.49379382
kerja gotong royong upaya putus rantai sebar virus covid jaga lonjak libur	0	0.011009037
pek indonesia suntik juta dosis vaksin masyarakat target anak usia vaksinasi capai hasil gotong royong	4	0.9463689
nu milik potensi rangka perata ekonomi umat kuat anak muda santri kualitas kompetensi rajut lokomotif tarik gerbong sejahtera	1	0.9547293
buka muktamar nahdlatul ulama lampung siang terima kasih nu awal jalan perintah senantiasa depan bangsa toleransi majemuk pancasila uud sea nkri	0	0.9609101

4. KESIMPULAN

Berdasarkan hasil pengujian dan analisis yang telah dilakukan, maka dapat disimpulkan bahwa model LDA dapat merepresentasikan topik topik yang ada di dalam dokumen tweets. Sehingga metode LDA dapat digunakan untuk mencari atau menemukan topik yang ada pada tweet para pejabat negara. Dengan jumlah topik yang dipilih sebanyak 7 dalam penelitian ini, diperoleh nilai perplexity sebesar -8.068 dan nilai Coherence Score sebesar 0.374. Namun, nilai coherence score tertinggi berada pada jumlah topik sebanyak 6. Hal tersebut menunjukkan bahwa pemilihan jumlah topik menjadi parameter dalam mencari nilai coherence yang terbaik. Penentuan jumlah topik yang sesuai akan menghasilkan nilai coherence scor yang valid dan optimal sehingga dapat diinterpretasikan. Dari hasil topik yang diperoleh dalam penelitian ini, dapat kita mengambil kesimpulan bahwa tweets yang dibagikan oleh Presiden RI Ir. Jokowi menggambarkan fokus perhatian presiden saat ini adalah kasus covid – 19 serta vaksinasi pada masyarakat Indonesia. Sebagai saran, perlu dilakukan analisis topik dengan menggunakan metode yang berbeda. Tujuannya, yaitu sebagai perbandingan dalam mencari model terbaik untuk menganalisis topik dari tweet yang dibagikan oleh para pejabat negara.

REFERENCES

[1] S. R. I. Rezeki, "Penggunaan Sosial Media Twitter dalam Komunikasi Organisasi (Studi Kasus Pemerintah Provinsi Dki Jakarta Dalam Penanganan Covid-19)," *J. Islam. Law Stud.*, vol. 04, no. 02, pp. 63–78, 2020.

[2] J. C. Campbell, A. Hindle, and E. Stroulia, "Latent Dirichlet Allocation: Extracting Topics from Software Engineering Data," *Art Sci. Anal. Softw. Data*, vol. 3, pp. 139–159, 2015, doi: 10.1016/B978-0-12-411519-4.00006-9.

[3] P. Zambrano *et al.*, "Technical mapping of the grooming anatomy using machine learning paradigms: An information security approach," *IEEE Access*, vol. 7, pp. 142129–142146, 2019, doi: 10.1109/ACCESS.2019.2942805.

[4] M. L. C. Chilmi, "Latent dirichlet allocation lda untuk mengetahui topik pembicaraan warganet twitter tentang omnibus law," *Repository.Uinjkt.Ac.Id*, 2021, [Online]. Available: https://repository.uinjkt.ac.id/dspace/handle/123456789/56724%0Ahttps://repository.uinjkt.ac.id/dspace/bitstream/123456789/56724/1/M. LUVIAN CHISNI CHILMI-FST.pdf.

[5] B. W. Arianto and G. Anuraga, "Topic Modeling for Twitter Users Regarding the 'Ruangguru' Application," *J. ILMU DASAR*, vol. 21, no. 2, p. 149, 2020, doi: 10.19184/jid.v21i2.17112.

[6] S. A. Putri, P. D. Kusuma, C. Setianingsih, and U. Telkom, "Clustering Topik Pada Data Sentimen Bpjs Kesehatan Menggunakan Metode Latent Dirichlet Allocation Topic Clustering On Sentiment Data Of Bpjs Kesehatan," vol. 8, no. 5, pp. 6097–6105, 2021.

[7] Z. Tong and H. Zhang, "A Text Mining Research Based on LDA Topic Modelling," pp. 201–210, 2016, doi: 10.5121/csit.2016.60616.

[8] Y. Sahria and D. H. Fudholi, "Analisis Topik Penelitian Kesehatan di Indonesia Menggunakan Metode Topic Modeling LDA," *J. Rekayasa Sist. dan Teknol. Inf.*, vol. 4, no. 2, pp. 336–344, 2020.

[9] D. Newman, J. H. Lau, K. Grieser, and T. Baldwin, "Automatic evaluation of topic coherence," *NAACL HLT 2010 - Hum. Lang. Technol. 2010 Annu. Conf. North Am. Chapter Assoc. Comput. Linguist. Proc. Main Conf.*, no. June, pp. 100–108, 2010.

[10] T. Gonçalves and P. Quaresma, "Evaluating preprocessing techniques in a text classification problem," *Unisinos*, pp. 841–850, 2005, [Online]. Available: http://www.research.att.com/.

[11] F. Rashif, G. Ihza Perwira Nirvana, M. Alif Noor, and N. Aini Rakhmawati, "Implementasi LDA untuk Pengelompokan Topik Cuitan Akun Bot Twitter bertagar #Covid-19 LDA Implementation for Topic of Bot's Tweets with #Covid-19 Hashtag," *Cogito Smart J.* |, vol. 7, no. 1, pp. 170–181, 2021.

[12] S. H. Mohammed and S. Al-Augby, "LSA & LDA topic modeling classification: Comparison study on E-books," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 19, no. 1, pp. 353–362, 2020, doi: 10.11591/ijeecs.v19.i1.pp353-362.

[13] A. I. Alfanzar and I. S. Rozas, "Topic Modelling Skripsi Menggunakan Metode Latent," vol. 7, no. 1, pp. 7–13, 2020.

[14] A. F. Hidayatullah and M. R. Ma'Arif, "Road traffic topic modeling on Twitter using latent dirichlet allocation," *Proc. - 2017 Int. Conf. Sustain. Inf. Eng. Technol. SIET 2017*, vol. 2018-Janua, no. August, pp. 47–52, 2018, doi: 10.1109/SIET.2017.8304107.

Building of Informatics, Technology and Science (BITS)
Volume 3, No 3, December 2021, Page 122−129
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v3i3.1012

[15] T. Santika, *Evaluasi Perplexity Untuk Pemodelan Topik Diskusi Agama Islam Di Media Sosial Twitter Indonesia Tahun 2006-2018 Menggunakan Latent Dirichlet Allocation Program Studi Matematika Uin Syarif Hidayatullah Jakarta*. 2019.
[16] H. S. Koh and M. Fienup, "Topic modeling as a tool for analyzing library chat transcripts," *Inf. Technol. Libr.*, vol. 40, no. 3, 2021, doi: 10.6017/ital.v40i3.13333.
[17] D. Mimno, H. M. Wallach, E. Talley, M. Leenders, and A. McCallum, "Optimizing semantic coherence in topic models," *EMNLP 2011 - Conf. Empir. Methods Nat. Lang. Process. Proc. Conf.*, no. 2, pp. 262–272, 2011.