

Analisis Topik Informasi Publik Media Sosial di Surabaya Menggunakan Pemodelan *Latent Dirichlet Allocation (LDA)*

I Made Kusnanta Bramantya Putra dan Renny Pradina Kusumawardani

Departemen Sistem Informasi, Fakultas Teknologi Informasi, Institut Teknologi Sepuluh Nopember (ITS)

e-mail: renny.pradina@gmail.com

Abstrak—Radio Suara Surabaya sebagai salah satu radio di Kota Surabaya merupakan radio yang menerapkan format "radio News" dan informasi, dimana informasi yang sering di sampaikan antara lain kondisi lalu lintas, keamanan, dan seputar Kota Surabaya. Radio Suara Surabaya mengembangkan siaran interaktif berbasis jurnalistik masyarakat, dimana agasan ini melibatkan partisipasi warga dalam melaporkan peristiwa kepada penyiar radio yang sedang bertugas. Laporan masyarakat yang masuk kemudian disebarluaskan kembali baik oleh penyiar melalui siaran radio maupun oleh *gatekeeper* melalui media social yang dimiliki Radio Suara Surabaya, baik *twitter* dan *Facebook* untuk memberikan informasi kepada masyarakat. Tingginya jumlah laporan perhari yang masuk melalui media social dan beragamnya topik dari laporan tersebut menimbulkan kesulitan dalam mengidentifikasi suatu topik dari kumpulan laporan media social masyarakat dan menghabiskan banyak waktu jika dilakukan secara manual oleh manusia. Padahal, kumpulan laporan tersebut merupakan sumber data yang sangat berpotensi untuk memberikan informasi apa yang terjadi di Kota Surabaya. Dengan kondisi demikian, dibutuhkan suatu pemodelan topik yang mampu secara otomatis mengklasifikasikan pesan media social ke dalam topik-topik yang muncul dari hasil pemodelan. Pemodelan topik dilakukan dengan metode *Latent Dirichlet Allocation (LDA)*, sebuah metode text mining untuk menemukan pola tertentu pada sebuah dokumen dengan menghasilkan beberapa macam topik yang berbeda. Eksperimen pemodelan *topic* dengan metode *LDA* menyimpulkan bahwa jumlah topik yang terdapat dalam pesan media social adalah 4 topik. Hasil eksperimen ini telah diuji secara mesin dengan nilai *perplexity* terbaik sebesar 213.41 dan diuji kemudahannya untuk diinterpretasi oleh manusia melalui uji koherensi topik yang terdiri dari *word intrusion task* dan *topic intrusion task*. Kesimpulan dari uji koherensi topik menyatakan bahwa model yang dihasilkan dengan metode *LDA* pada studi kasus ini dapat diinterpretasi manusia dengan baik.

Kata Kunci—Radio Suara Surabaya, pesan media sosial, *Latent Dirichlet Allocation (LDA)*, uji koherensi topik, *word intrusion task*, *topic intrusion task*.

I. PENDAHULUAN

KOTA Surabaya merupakan ibu kota Provinsi Jawa Timur Indonesia, sekaligus menjadi kota metropolitan terbesar di provinsi tersebut. Surabaya sebagai kota terbesar kedua di Indonesia setelah Jakarta juga merupakan pusat bisnis, perdagangan, industri, dan pendidikan di Jawa Timur serta wilayah Indonesia lainnya di bagian timur. Secara geografis, Kota Surabaya terletak di 07° 15' 50" Lintang Selatan dan 112° 45' 03" Bujur Timur, atau di tepi pantai utara Pulau Jawa dan berhadapan dengan Selat Madura serta Laut Jawa.

Surabaya memiliki luas sekitar 333,063 km² dengan penduduknya berjumlah 2.909.257 jiwa (2015) [1].

Salah satu keunikan Kota Surabaya adalah adanya Radio Suara Surabaya FM (SSFM) yang memiliki slogan *News, Interaktif dan Solutif*, sehingga dalam operasionalnya, tidak menyiarkan musik dan hiburan seperti radio lainnya, namun SSFM ini berhasil menjadi penghubung masyarakat yang memonitor kondisi keseharian warga Kota Surabaya dan sekitarnya. SSFM mampu meningkatkan partisipasi pendengarnya sebagai "reporter jalanan" dalam bentuk siaran interaktif dengan penyiar dan *gatekeeper*. Konsistensi SSFM sebagai radio berita interaktif telah mampu menjadi lembaga yang mawadahi interaksi sosial sehari-hari dalam masyarakat [2].

Tidak hanya melalui radio, Radio Suara Surabaya FM juga mengikuti perkembangan jaman di masyarakat dengan merambah media sosial. Akun *e100ss* adalah akun *Twitter* dan *Facebook* resmi Radio Suara Surabaya yang dimanfaatkan sebagai salah satu saluran media komunikasi untuk berbagi berita secara *real-time* dengan pendengar untuk mendukung siaran Suara Surabaya FM (SSFM) [3].

Jumlah informasi publik media sosial yang di-*post* melalui media sosial resmi Radio Suara Surabaya FM sangat tinggi, hal ini ditunjukkan dengan jumlah rata-rata *tweet* dalam sehari dari akun *@e100ss* sejumlah 192,2, sementara jumlah *post* dalam sehari dari akun *E100* sejumlah 126,18 *post*, yang mana kedua data tersebut belum termasuk pesan dari pendengar atau pengguna yang ditujukan kepada media sosial resmi Radio Suara Surabaya FM. Dengan jumlah *post* yang tinggi, terdapat pula hal lain yang perlu disoroti, yaitu pola pesan yang tidak sesuai dengan ejaan yang disempurnakan dan topik yang beragam dan sangat cepat berubah dari satu waktu ke waktu, sehingga untuk mengidentifikasi suatu topik dari kumpulan *tweet* tersebut sangat sulit dan menghabiskan banyak waktu jika dilakukan secara manual oleh manusia. Padahal, kumpulan *post* tersebut merupakan sumber data yang sangat berpotensi untuk memberikan informasi apa yang terjadi di Kota Surabaya.

Berdasarkan permasalahan tersebut, penelitian ini akan menawarkan solusi dalam melakukan analisis *topic modelling* terhadap pesan media sosial yang disampaikan melalui akun *e100ss*. Analisis *topic modelling* digunakan untuk mengetahui topik-topik apa saja yang sering muncul, sehingga memudahkan dalam mengetahui apa yang sedang terjadi di kota Surabaya. Penelitian ini menggunakan pembelajaran *machine learning* dengan metode *Latent Dirichlet Allocation (LDA)* [4]. Metode *LDA* merupakan sebuah metode *text mining* untuk menemukan pola tertentu pada sebuah dokumen dengan menghasilkan beberapa macam topik yang berbeda, sehingga tidak secara spesifik mengelompokkan dokumen

kedalam sebuah topik tertentu [4]. Topik yang muncul dari pengolahan data tersebut selanjutnya akan dilakukan uji koherensi topik, yaitu uji keterkaitan dari uraian probabilitas kata-kata yang ditemukan satu sama lain dalam menyusun suatu topik [5].

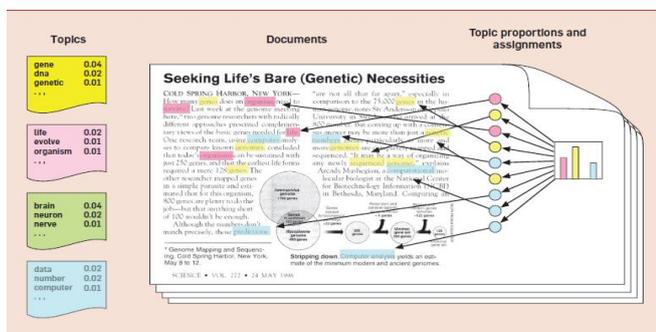
Hasil dari analisis *topic modelling* menggunakan *LDA* diharapkan dapat membantu dalam memahami apa yang sedang terjadi di Surabaya secara lebih ringkas.

II. DASAR TEORI

A. Topic modelling

Konsep *topic modelling* menurut Blei terdiri dari entitas-entitas yaitu “kata”, “dokumen”, dan “*corpora*”. “Kata” dianggap sebagai unit dasar dari data diskrit dalam dokumen, didefinisikan sebagai item dari kosa kata yang diberi indeks untuk setiap kata unik pada dokumen. “Dokumen” adalah susunan N kata-kata. Sebuah *corpus* adalah kumpulan M dokumen dan *corpora* merupakan bentuk jamak dari *corpus*. Sementara “*topic*” adalah distribusi dari beberapa kosakata yang bersifat tetap. Secara sederhana, setiap dokumen dalam *corpus* mengandung proporsi tersendiri dari topik-topik yang dibahas sesuai kata-kata yang terkandung di dalamnya [4].

Ide dasar dari *topic modeling* adalah bahwa sebuah topik terdiri dari kata-kata tertentu yang menyusun topik tersebut, dan dalam satu dokumen memiliki kemungkinan terdiri dari beberapa topik dengan probabilitas masing-masing. Namun secara pemahaman manusia, dokumen-dokumen merupakan objek yang dapat diamati, sedangkan topik, distribusi topik per-dokumen, dan penggolongan setiap kata pada topik per-dokumen merupakan struktur tersembunyi, maka dari itu *topic modelling* bertujuan untuk menemukan topik dan kata-kata yang terdapat pada topik tersebut [6]. Konsep *topic modelling* menurut Blei, ditunjukkan pada **Gambar 1**



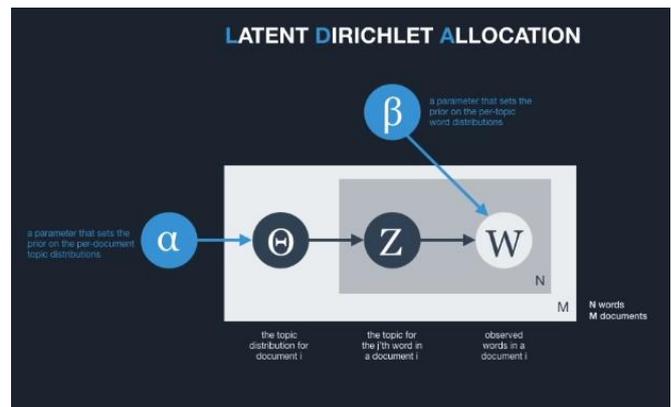
Gambar 1. Latent Dirichlet Allocation (LDA) menurut Blei

B. Latent Dirichlet Allocation (LDA)

Latent Dirichlet Allocation (LDA) merupakan metode topik modelling dan topik analisis yang paling populer saat ini. *LDA* muncul sebagai salah satu metode yang dipilih dalam melakukan analisis pada dokumen yang berukuran sangat besar. *LDA* dapat digunakan untuk meringkas, melakukan klusterisasi, menghubungkan maupun memproses data yang sangat besar karena *LDA* menghasilkan daftar topik yang diberi bobot untuk masing-masing dokumen [7]. Adapun distribusi yang digunakan untuk mendapatkan distribusi topik per-dokumen disebut distribusi Dirichlet, kemudian dalam proses generatif untuk *LDA*, hasil dari Dirichlet digunakan untuk mengalokasikan kata-kata pada dokumen untuk topik yang berbeda. Dalam *LDA*, dokumen-dokumen merupakan objek yang dapat diamati, sedangkan topik, distribusi topik

per-dokumen, penggolongan setiap kata pada topik per-dokumen merupakan struktur tersembunyi, maka dari itu, Algoritma ini dinamakan *Latent Dirichlet Allocation (LDA)* [6].

Menurut Blei (2003), *LDA* merupakan model probabilistik generatif dari kumpulan tulisan yang disebut *corpus*. Ide dasar yang diusulkan metode *LDA* adalah setiap dokumen direpresentasikan sebagai campuran acak atas topik yang tersembunyi, yang mana setiap topik memiliki karakter yang ditentukan berdasarkan distribusi kata-kata yang terdapat di dalamnya [4]. Blei merepresentasikan metode *LDA* sebagai model *probabilistic* secara visual seperti pada **Gambar 2**.



Gambar 2. Visualisasi Topic modelling dengan Metode LDA [8]

Sesuai visualisasi model di atas, terdapat tiga tingkatan pada *LDA* Modeling. Parameter α dan β merupakan parameter distribusi topik yang berada pada tingkatan *corpus*, yaitu kumpulan dari M dokumen. Parameter α digunakan dalam menentukan distribusi topik dalam dokumen, semakin besar nilai alpha dalam suatu dokumen, menandakan campuran topik yang dibahas dalam dokumen semakin banyak. Parameter β digunakan untuk menentukan distribusi kata dalam topik. Semakin tinggi nilai beta, maka semakin banyak kata-kata yang ada di dalam topik, sedangkan semakin kecil nilai beta, maka semakin sedikit kata-kata yang ada di dalam topik sehingga topik tersebut mengandung kata-kata yang lebih spesifik. Variabel θ_m adalah variabel yang berada di tingkat dokumen (M). Variabel θ merepresentasikan distribusi topik untuk dokumen tertentu. Semakin tinggi nilai θ , maka semakin banyak topik yang ada di dalam dokumen, sedangkan semakin kecil nilai θ , maka dapat dikatakan dokumen tersebut semakin spesifik pada topik tertentu. Variabel Z_n dan W_n adalah variabel tingkat kata (N). Variabel Z dan merepresentasikan topik dari kata tertentu pada sebuah dokumen sedangkan variabel W merepresentasikan kata yang berkaitan dengan topik tertentu yang terdapat dalam dokumen [4].

Secara umum, *LDA* bekerja dengan masukan dokumen-dokumen individual dan beberapa parameter, untuk menghasilkan luaran berupa model yang terdiri dari bobot yang dapat dinormalisasi sesuai probabilitas. Probabilitas ini mengacu pada dua jenis, yaitu jenis (a) probabilitas bahwa suatu dokumen spesifik tertentu menghasilkan topik yang spesifik pula dan jenis (b) probabilitas bahwa topik spesifik tertentu menghasilkan kata-kata spesifik dari sebuah kumpulan kosakata. Probabilitas jenis (a), dokumen yang sudah diberi label dengan daftar topik seringkali dilanjutkan hingga menghasilkan probabilitas jenis (b), yang menghasilkan kata-kata spesifik tertentu [7].

LDA dapat digunakan dalam bidang analisis trend pada media social [8], melakukan identifikasi tema pada 421 artikel ilmiah pada *Organizational Research Methods* [9], mendeteksi topik untuk pelacakan konten percakapan [10] dan telah terbukti mampu berkerja dengan baik untuk dokumen panjang seperti artikel *Wikipedia* maupun dokumen pendek seperti *tweet* [11]

C. Radio Suara Surabaya

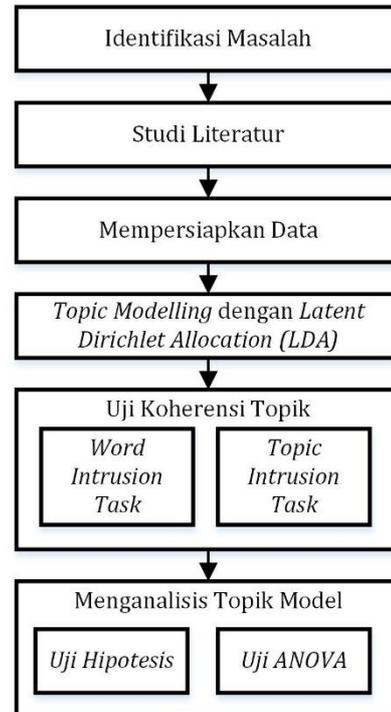
Radio Suara Surabaya Suara Surabaya FM (SSFM) adalah sebuah stasiun radio di Kota Surabaya. Berpusat di jalan Wonokitri Besar 40, SSFM mengudara pertama kali pada tanggal 11 Juni 1983. Radio ini mengklaim sebagai radio pertama di Indonesia yang menerapkan format "radio News" dan informasi, dimana informasi yang sering di sampaikan antara lain kondisi lalu lintas, keamanan, dan seputar Kota Surabaya [12].

Pada tahun 1994, Radio SS mengembangkan siaran interaktif berbasis jurnalistik masyarakat. Gagasan ini melibatkan partisipasi warga dalam melaporkan peristiwa kepada penyiar radio yang sedang bertugas [13]. Partisipasi yang melibatkan warga ini diawali berdasarkan inisiatif masyarakat sendiri, yang melaporkan kondisi lalu lintas dan cuaca yang ditanggapi positif oleh SSFM dengan menyebarkan laporan tersebut sehingga pendengar merasakan langsung manfaatnya. Menariknya SS tidak pernah mengadakan training atau menentukan format tertentu untuk pelaporan yang dilakukan masyarakat. SS memiliki penyiar radio dan gatekeeper yang memiliki peran untuk menyeleksi atau menyunting informasi yang disampaikan oleh masyarakat melalui media sosial [14].

Akun *Twitter* @e100ss dan akun *Facebook* E100 adalah akun media sosial resmi Radio Suara Surabaya sebagai salah satu saluran media komunikasi yang mendukung siaran Suara Surabaya FM (SSFM). Akun @e100ss dikembangkan Tim *On Air* Suara Surabaya pada September 2012. Nama @e100ss dipilih dengan tujuan agar mudah diingat, e menandakan *electronic*, 100 adalah Frekuensi Suara Surabaya FM dan ss merupakan singkatan dari Suara Surabaya. Sesuai visi SSFM yg mengikuti perkembangan teknologi, Suara Surabaya FM memanfaatkan media sosial @e100ss dan E100 sebagai media komunikasi resmi dengan pendengar untuk berbagi berita secara *real-time* [3].

III. METODOLOGI

Langkah-langkah pengerjaan studi dari awal hingga akhir ditunjukkan pada Gambar 3.



Gambar 3. Metodologi Penelitian

IV. HASIL DAN ANALISIS

Dalam melakukan *topic modelling* dengan *Latent Dirichlet Allocation (LDA)*, tahap-tahap yang dilakukan adalah preprocessing data, analisis jumlah iterasi dan analisis jumlah topik

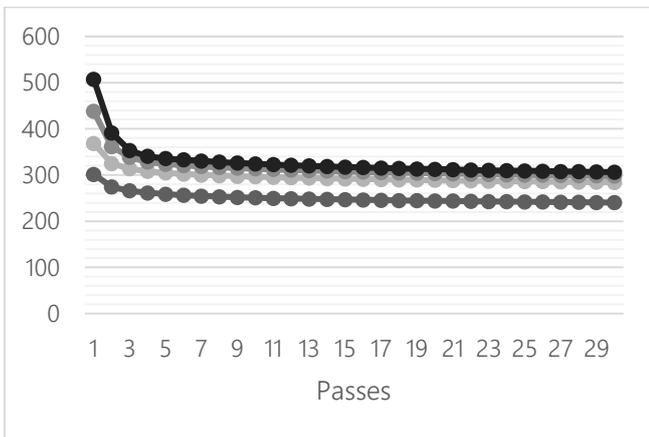
A. Preprocessing Data

Pra-proses data mencakup beberapa langkah utama pengerjaan yakni membersihkan data dan *stopword* removal. Pembersihan data dilakukan untuk menghilangkan penulisan huruf besar menjadi huruf kecil, menghilangkan adanya karakter yang tidak diperlukan, menghapus kata dengan jumlah huruf kurang dari 2 dan menghapus token yang hanya terdiri dari angka.

Pendefinisian *stopword* mengacu pada susunan *stopword* yang telah disusun oleh penelitian Fadillah Z Talla [15] yang dilengkapi dengan pendefinisian kata tugas berdasarkan buku karangan Masnur Muslich [16] dan Abdul Muslich [17], kata tugas dipilih menjadi *stopword* karena merupakan kata yang sering digunakan namun tidak memiliki nilai informasi. Selain itu, terdapat tambahan kata-kata yang menjadi kode format dari Radio Suara Surabaya FM. Kemudian hasil dari kedua tahap ini, data diterjemahkan untuk menjadi *corpus* serta disimpan untuk penggunaan pemrosesan data selanjutnya.

B. Analisis Jumlah Iterasi

Dalam melakukan penentuan jumlah iterasi (*passes*), metode yang digunakan adalah dengan melakukan analisis nilai *perplexity*. Analisis nilai *perplexity* dilakukan dengan cara menjalankan pemodelan terhadap topik dengan setidaknya tiga parameter topik yang berbeda dengan nilai *passes* mula-mula 30. Dalam kasus ini, parameter jumlah topik yang dipilih adalah 5, 10, 15 dan 20. Hasil nilai *perplexity* yang muncul dari masing-masing parameter jumlah topik ini kemudian dicatat kemudian divisualisasikan pada Gambar 4

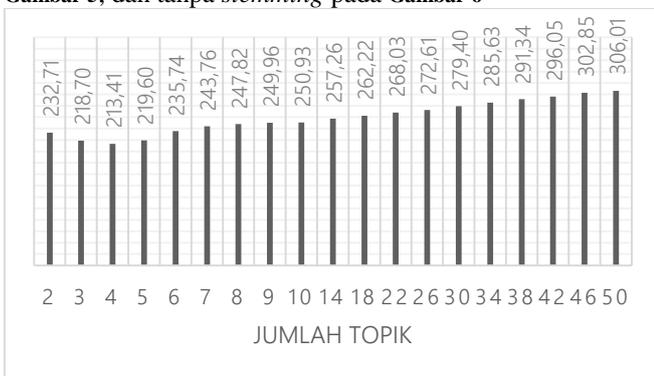


Gambar 4. Visualisasi hasil eksperimen untuk penentuan jumlah iterasi dengan analisis *perplexity*

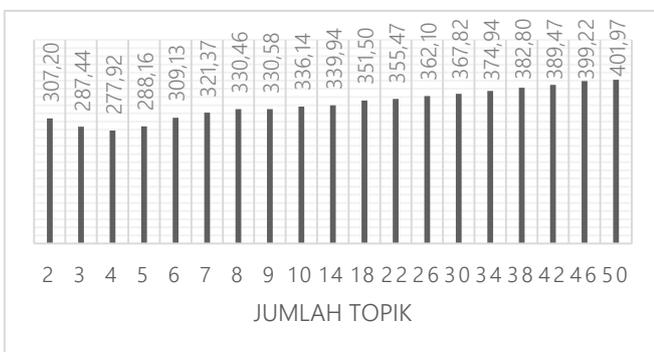
Berdasarkan visualisasi Gambar 4, dapat dilihat bahwa nilai *perplexity* sudah mencapai kondisi cenderung stabil pada *passes* ke 10 untuk keseluruhan parameter jumlah topik, sehingga dapat disimpulkan bahwa iterasi yang digunakan adalah 10.

C. Analisis Jumlah Topik

Dalam melakukan eksperimen untuk penentuan jumlah topik dengan analisis *perplexity*, eksperimen dibedakan menjadi eksperimen pemodelan topik dengan stem dan tanpa stem untuk mengetahui pengaruh stem dalam pemodelan topik. Hasil eksperimen analisis nilai *perplexity* untuk penentuan jumlah topik dengan *stemming* dapat dilihat pada Gambar 5, dan tanpa *stemming* pada Gambar 6



Gambar 5. Visualisasi hasil eksperimen untuk penentuan jumlah topik dengan analisis *perplexity* untuk model dengan stem



Gambar 6. Visualisasi hasil eksperimen untuk penentuan jumlah topik dengan analisis *perplexity* untuk model tanpa stem

Berdasarkan Gambar 5, nilai *perplexity* terendah untuk eksperimen *topic modeling* dengan stem terdapat pada jumlah topik 4 yaitu 213.41, sementara berdasarkan Gambar 6 untuk

eksperimen topik modeling tanpa stem nilai *perplexity* terendah terdapat pada jumlah topik 4 yaitu 277.92. Tren nilai *perplexity* meningkat untuk jumlah topik yang semakin tinggi, sehingga 4 topik merupakan jumlah topik terbaik berdasarkan analisis *perplexity*.

V. KESIMPULAN

Kesimpulan yang didapatkan dari proses pengerjaan studi yang telah dilakukan antara lain:

1. Dalam melakukan *topic modelling* dengan metode *LDA* untuk menganalisis topik-topik apa yang sedang dibahas pada media sosial informasi publik di Surabaya, tahap tahap yang dilakukan adalah mengidentifikasi masalah, melakukan studi literature, mempersiapkan data, melakukan pra-proses data, membentuk model topik, serta melakukan validasi model topik
2. Telah dilakukan berbagai eksperimen dalam proses *topic modelling* dengan metode *LDA*. Eksperimen yang terkait jumlah *passes* menghasilkan 10 *passes* sebagai jumlah *passes* terbaik berdasarkan nilai *perplexity* yang telah stabil pada iterasi ke 10 untuk setiap distribusi topik yang terpilih secara acak.
3. Eksperimen yang terkait dengan jumlah topik dalam model dibagi menjadi dua eksperimen, yaitu *topic modelling* dengan stem dan *topic modelling* tanpa stem. *Topic modelling* dengan stem menghasilkan kesimpulan bahwa 4 topik merupakan jumlah topik terbaik berdasarkan nilai *perplexity* terendah yaitu 213.41 dengan standar deviasi 2.54. *Topic modelling* tanpa stem juga menghasilkan kesimpulan bahwa 4 topik merupakan jumlah topik terbaik berdasarkan nilai *perplexity* terendah yaitu 277.92 dengan standar deviasi 3.67. Berdasarkan hasil eksperimen ini, telah dapat dihasilkan beberapa kesimpulan. Kesimpulan pertama, bahwa 4 topik merupakan jumlah topik terbaik dalam membentuk topik model berdasarkan analisis nilai *perplexity* dan dapat dikatakan cukup stabil berdasarkan nilai standar deviasinya. Kesimpulan kedua, proses *stemming* dalam kasus ini memberikan pengaruh positif terhadap pembentukan model yang ditunjukkan dengan penurunan nilai *perplexity* sebesar 64.51.
4. Eksperimen terkait klasifikasi dokumen ke dalam topik telah dilakukan baik untuk model dengan stem, maupun model tanpa stem. Hasil eksperimen klasifikasi dokumen menunjukkan hasil yang meyakinkan yang ditunjukkan dengan distribusi probabilitas dokumen terhadap topik untuk model dengan atau tanpa stem secara mayoritas berada pada rentang 0.5 hingga 0.99.

VI. SARAN

Dari pengerjaan studi ini, terdapat beberapa saran untuk pengembangan penelitian ke depan.

1. Data yang digunakan dalam penelitian ini bersumber dari status *Facebook* fanpage dan *twitter* yang menggunakan kebahasaan berita media sosial. Pada pengembangannya, dapat diujicobakan dataset dari akun media social lain yang terpercaya ataupun dataset yang diambil dari komentar pada status *Facebook* fanpage yang cenderung lebih banyak menggunakan karakteristik bahasa informal sehari-hari.

2. Dari hasil pemodelan topik, terdapat kata-kata yang tergolong sebagai nama lokasi dalam topik, sehingga untuk memperoleh hasil yang lebih optimal, diperlukan suatu pendeteksian nama lokasi sebagai penyaringan data sebelum dilakukan pemodelan, demikian pula untuk merk dagang dan sebagainya. Terdapat pula kata-kata yang berupa singkatan yang memiliki arti sama namun dalam pemodelan topik menjadi dua entitas yang berbeda, seperti jln dengan jalan dan yg dengan yang. Untuk itu diperlukan suatu normalisasi sebelum dilakukan pemodelan topik

LAMPIRAN

Tabel 1.
Hasil Pembentukan Model LDA dengan Stemming

topik #0:	topik #1:	topik #2:	topik #3:
0.021*surabaya	0.017*hujan	0.035*info	0.073*arah
0.010*kawan	0.016*surabaya	0.023*celaka	0.069*padat
0.008*indonesia	0.015*pagi	0.019*lokasi	0.044*macet
0.007*suara	0.014*kawan	0.015*motor	0.030*jalur
0.006*selamat	0.014*jalan	0.015*truk	0.027*tol
0.005*news	0.011*jl	0.015*foto	0.024*imbas
0.005*anak	0.011*banjir	0.015*kendara	0.021*via
0.005*wib	0.010*selamat	0.014*via	0.020*foto
0.005*foto	0.010*kawasan	0.014*mobil	0.017*truk
0.005*orang	0.009*air	0.013*tugas	0.016*waru
0.004*radio	0.009*cuaca	0.013*lapor	0.012*lintas
0.004*ibadah	0.008*wib	0.012*korban	0.012*simpang
0.004*ciri	0.007*kota	0.012*libat	0.012*surabaya
0.004*hilang	0.006*deras	0.010*jalan	0.012*jalan
0.004*kumandang	0.006*daerah	0.010*bakar	0.012*lajur
0.004*adzan	0.006*gubernur	0.010*update	0.010*mogok
0.004*tunai	0.005*hati	0.010*sepeda	0.010*waspada
0.004*jatim	0.005*listrik	0.009*data	0.010*arus
0.003*tiket	0.005*kerja	0.009*kronologi	0.009*raya
0.003*rumah	0.005*suhu	0.008*dengar	0.009*sidoarjo

Tabel 2.
Hasil Pembentukan Model LDA tanpa Stemming

topik #0:	topik #1:	topik #2:	topik #3:
0.014*surabaya	0.030*info	0.070*padat	0.021*surabaya
0.008*indonesia	0.024*truk	0.068*arah	0.015*kawan
0.007*orang	0.021*kecelakaan	0.043*macet	0.014*hujan
0.006*suara	0.019*via	0.030*jalur	0.014*wib
0.006*news	0.018*foto	0.025*imbas	0.012*pagi
0.006*anak	0.017*lokasi	0.024*tol	0.011*selamat
0.006*foto	0.013*mobil	0.017*via	0.008*banjir
0.005*rumah	0.013*motor	0.017*surabaya	0.008*cuaca
0.005*kawan	0.013*jalan	0.016*foto	0.007*kawasan
0.004*ciri	0.013*petugas	0.015*waru	0.007*radio
0.003*jatim	0.011*update	0.012*simpang	0.007*jl

0.003*warga	0.010*km	0.011*jalan	0.007*tiket
0.003*presiden	0.009*korban	0.011*mogok	0.006*air
0.003*keluarga	0.009*melaporkan	0.009*raya	0.006*pk1
0.003*pk	0.009*kendaraan	0.009*truk	0.006*kota
0.003*tim	0.009*sepeda	0.009*kawan	0.006*rp
0.003*jawa	0.008*tol	0.009*arus	0.005*jalan
0.003*gk	0.008*arah	0.009*volume	0.005*deras
0.002*reporter	0.008*melibatkan	0.008*sidoarjo	0.005*wilayah
0.002*kepala	0.008*sidoarjo	0.008*exit	0.005*pk

DAFTAR PUSTAKA

[1] P. K. Surabaya, "Situs Web Resmi Kota Surabaya," 2016. [Online]. Available: <http://sparkling.surabaya.go.id/>.

[2] D. I. P. Hadi, "Penyiaran Interaktif dan Kepentingan Publik," 2014.

[3] R. S. Surabaya, "Suara Surabaya Official Website," 2016. [Online]. Available: <http://www.suarasurabaya.net/>.

[4] D. M. Blei, *Latent Dirichlet Allocation*. Machine Learning Research 3, 2003.

[5] J. H. L. K. G. T. B. D. Newman, "Automatic Evaluation of Topic Coherence," in *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL HLT 2011)*, 2010, pp. 100–108.

[6] D. M. Blei, "Probabilistic Topic Model," vol. 55, 2012.

[7] A. H. and a. E. S. J. C. Campbell, "Latent Dirichlet Allocation: Extracting Topics," 2014.

[8] N. C. T. B. J. H. Lau, "On-line Trend Analysis with Topic Models: #twitter trends detection topic model online," in *Proceedings of COLING 2012: Technical Papers*, 2012, pp. 1519–1534.

[9] A. Gaur, "Topic Models As A Novel Approach To Identify Themes In Content Analysis: The Example Of Organizational Research Methods," 2015.

[10] Y.-S. T. and C.-H. L. J.-F. Yeh, "Topic detection and tracking for conversational content by using conceptual dynamic latent Dirichlet Allocation," 2016.

[11] H. Z. Z. Tong, "A Text Mining Research Based On LDA Topic Modelling," *Jodrey Sch. Comput. Sci. Acadia Univ.*, vol. 10, pp. 201–210, 2016.

[12] H. T. Jukangko, "Peran Radio Suara Surabaya Sebagai Media Pendidikan untuk Tertib Berlalu Lintas bagi Warga Kota Surabaya," vol. 3, no. 1, 2013.

[13] I. P. Hadi, "Khalayak Maya Dalam Media Online," *J. Ilm. SCRIPTURA*, vol. 1, 2007.

[14] F. L. D. Yoanita, "Akurasi dalam Jurnalisma Warga pada Radio Suara Surabaya," *J. SCRIPTURA*, vol. 4, pp. 47–53, 2014.

[15] F. Z. Tala, "A Study of Stemming Effects on Information Retrieval in Bahasa Indonesia," *Inst. Logic, Lang. Comput. Univ. van Amsterdam, Netherlands*, 2003.

[16] M. Muslich, *Garis-garis Besar Tata Bahasa Baku Bahasa Indonesia*. Bandung: Refika Aditama, 2010.

[17] A. Chaer, *Tata Bahasa Praktis Bahasa Indonesia*. Jakarta: Barata Karya Aksara, 2008.