



Analisis *Topic Modelling* pada Ulasan Aplikasi Shopee di PlayStore Menggunakan *Latent Dirichlet Allocation (LDA)*

Sahat Renaldi. S¹, Primus Akbar Atnanda², Amri Muhaimin³

^{1,2,3} Program Studi Sains Data, UPN "Veteran" Jawa Timur

²20083010026@student.upnjatim.ac.id

³amri.muhamin.stat@upnjatim.ac.id

Corresponding author email: 20083010026@student.upnjatim.ac.id

Abstract: *Shopee is a popular e-commerce platform in Indonesia. To improve service quality, data analysis is needed to understand user responses and preferences. This study aims to conduct data analysis and topic modelling using the Latent Dirichlet Allocation (LDA) method on 33,896 user review data of the Shopee application on PlayStore. LDA modelling was performed by considering the parameters passes and iterations. Passes with values of 5, 10, 15, and 20 were tested with a combination of iterations values of 50 and 100 and a random state of 142. During the testing, it was found that the highest average coherence score was achieved by using passes 10 and iterations 50. Therefore, passes 10 and iterations 50 were selected as the final LDA model parameters, resulting in seven topics. The topics that emerged from the analysis include user satisfaction, shipping and product suitability, user trust, user experience, service and support, application, and promotional offers. The "application" topic was found to be the most critical and received the highest number of rating 1. High-weighted words on the application topic such as "slow," "heavy," and "crash" indicate that issues related to the application's system performance and functionality need to be addressed and improved.*

Keywords: *Shopee, Modelling, Latent Dirichlet Allocation*

Abstrak: *Shopee merupakan salah satu platform e-commerce yang populer di Indonesia. Dalam upaya untuk meningkatkan kualitas layanan, analisis data perlu dilakukan untuk memahami respon dan preferensi pengguna terhadap Shopee. Penelitian ini bertujuan untuk melakukan analisis data dan pemodelan topik dengan metode Latent Dirichlet Allocation (LDA) pada 33.896 data ulasan pengguna aplikasi Shopee di PlayStore. Pemodelan LDA dilakukan dengan mempertimbangkan parameter passes dan iterations. Parameter passes dengan nilai 5, 10, 15, 20 diuji dengan kombinasi nilai iterations 50 dan 100 serta random state 142. Dalam pengujian, ditemukan bahwa nilai rata-rata coherence score tertinggi terdapat pada kombinasi nilai passes 10 dan iterations 50. Nilai passes 10 dan iterations 50 dipilih sebagai parameter model LDA final dengan jumlah topik sebanyak tujuh topik. Adapun topik-topik yang terbantu antara lain topik kepuasan pengguna, topik pengiriman dan kesesuaian barang, topik kepercayaan pengguna, topik pengalaman pengguna, topik layanan dan bantuan, topik aplikasi, dan topik terkait promosi dan penawaran. Topik "aplikasi" menjadi topik yang paling kritis dan mendapatkan rating 1 terbanyak. Kata-kata dengan bobot tinggi pada topik aplikasi seperti "lambat", "berat", dan "buka" menunjukkan bahwa masalah terkait kinerja sistem dan fungsionalitas aplikasi menjadi salah satu isu yang perlu dibenahi.*

Kata kunci: *Shopee, Pemodelan, Latent Dirichlet Allocation*

I. PENDAHULUAN

Belanja online telah menjadi gaya hidup banyak orang di zaman digital sekarang ini. Berbicara tentang belanja online, Shopee merupakan salah satu platform belanja online terkemuka di Indonesia dengan jumlah pengguna aktif yang terus meningkat setiap tahunnya. Berdasarkan hasil survei dalam laporan Status Literasi Digital Indonesia tahun 2021 yang dikeluarkan oleh Kementerian Komunikasi dan Informatika dan Katadata Insight Center (KIC), Shopee merupakan *e-commerce* yang paling banyak digunakan masyarakat di Indonesia [1]. Shopee menyediakan berbagai layanan yang memudahkan pengguna dalam melakukan transaksi pembayaran, pengiriman, maupun pengembalian barang. Shopee juga menawarkan promo menarik dan program loyalitas, seperti melalui *flash sale*, voucher gratis ongkir, *Shopee coins*, dan program lainnya. Hal ini membuat Shopee semakin diminati oleh pengguna dan memperkuat posisi Shopee sebagai salah satu platform belanja online di Indonesia.



Dalam lingkup persaingan bisnis yang semakin kompetitif, Shopee harus selalu memperhatikan kebutuhan dan preferensi pengguna. Ulasan pengguna Shopee yang ada di PlayStore dapat menjadi salah satu sumber informasi bagi perusahaan untuk mengevaluasi kualitas layanan yang diberikan. Diharapkan dari data ulasan yang ada dapat menjadi bahan pembelajaran dan peningkatan pelayanan bagi pengguna. Salah satu cara yang dapat digunakan untuk menganalisis data teks ulasan pengguna adalah dengan *topic modelling*. *Topic modelling* adalah metode yang dapat digunakan untuk mengekstrak topik atau tema dari kumpulan dokumen teks yang besar dan kompleks. Dalam konteks ulasan pengguna Shopee di PlayStore, *topic modelling* dapat digunakan untuk mengidentifikasi topik atau tema yang sering dibicarakan oleh pengguna. Salah satu metode *topic modelling* yang umum digunakan adalah *Latent Dirichlet Allocation (LDA)*. LDA mengasumsikan bahwa setiap dokumen terdiri dari beberapa topik yang berbeda dan setiap kata dalam dokumen tersebut berasal dari salah satu topik tersebut

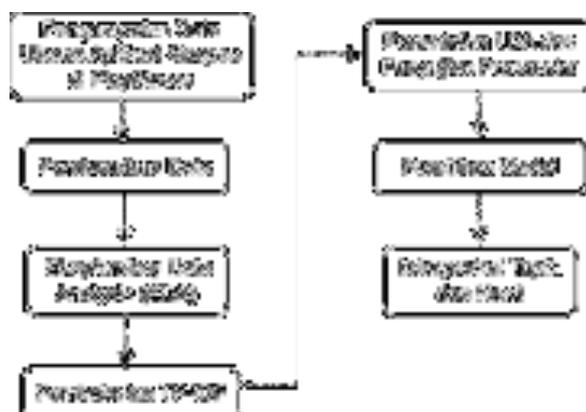
Beberapa penelitian terdahulu terkait *topic modelling* yang menggunakan LDA antara lain penelitian yang dilakukan oleh Habibi et al. (2021) menganalisis *caption* konten terkait Gerakan Masyarakat Hidup Sehat (GERMAS) di Instagram. Berdasarkan hasil analisis konten pada masing-masing segmen topik, ditemukan bahwa topik yang paling dominan terkait dengan GERMAS adalah pola makan hidup sehat [2]. Penelitian yang dilakukan oleh Destarani et al. (2019) memodelkan data keluhan yang dilaporkan melalui situs web keluhan online masyarakat Denpasar menggunakan LDA dengan estimasi parameter Gibbs *sampling*. Jumlah topik yang diperoleh melalui nilai log-likelihood tertinggi adalah sebesar -42.528,84, nilai tersebut berada pada jumlah topik 19. Topik yang sedang tren didasarkan pada probabilitas topik tertinggi, yakni nilai probabilitas topik 0,055 yang dapat diinterpretasikan bahwa banyak penduduk Denpasar mengeluhkan tentang jalan rusak dan meminta untuk memperbaiki jalan [3]. Sutherland et al. (2020) meneliti 104.161 data ulasan online dari pelanggan akomodasi di Korea. Dengan menggunakan model LDA, penelitian ini berhasil mengekstrak dan memvalidasi topik-topik penting yang menjadi perhatian tamu. Implikasi dari temuan ini dapat membantu manajer dalam meningkatkan kepuasan pelanggan [4].

Berdasarkan hasil tinjauan pada penelitian-penelitian sebelumnya, metode *Latent Dirichlet Allocation (LDA)* mampu memodelkan data teks ke dalam topik-topik tertentu. Oleh karena itu, peneliti tertarik untuk menggunakan LDA dalam memodelkan data ulasan pengguna aplikasi Shopee di PlayStore. Penelitian ini menggunakan data ulasan yang terhitung sejak tanggal 1 April 2023 sampai 20 April 2023. Setelah dilakukan pembersihan pada data mentah, didapatkan 33.896 record data yang akan digunakan dalam pemodelan. Penelitian ini mempertimbangkan parameter *passes* dan *iterations* dalam pemodelan dengan random state 142. Untuk memilih *passes* dan *iteration* final yang akan digunakan, terlebih dahulu dilakukan pengujian dengan mengkombinasikan nilai *passes* dan *iterations*. Jumlah topik (k) yang akan digunakan dalam model akhir dilihat K-Optimum yang paling sering muncul dari hasil kombinasi parameter. Parameter *passes* 5, 10, 15, 20 akan dikombinasikan dan diuji dengan *iterations* 50, dan 100. Parameter *passes* dan *iterations* yang dipilih sebagai model final adalah parameter dengan rata-rata *coherence score* tertinggi. Setelah dilakukan pemodelan dengan metode LDA, topik-topik yang dihasilkan akan dianalisis dan diinterpretasikan. Hasil analisis tersebut diharapkan dapat memberikan wawasan yang berguna bagi Shopee dalam meningkatkan kualitas layanan dan produk.”

II. METODE PENELITIAN

Dalam penelitian ini, alur penelitian dimulai dari proses pengumpulan data, pembersihan data, Ekploratory Data Analysis (EDA), pembobotan dengan TF-IDF, pemodelan LDA dan pengujian

parameter, pemilihan model, interpretasi topik dan hasil.



Gambar 1 Alur Penelitian

3.1 Pengumpulan Data

Data yang digunakan dalam penelitian ini didapatkan dari hasil proses crawling data PlayStore pada ulasan aplikasi Shopee dengan menggunakan library *google_play_scraper* yang ada di Python. Data ulasan yang diambil terhitung sejak tanggal 1 April 2023 sampai 20 April 2023.

3.2 Pembersihan Data

Data teks ulasan aplikasi Shopee yang diperoleh dari proses crawling merupakan data mentah yang masih mengandung unsur-unsur yang tidak diperlukan. Pembersihan data dilakukan sebagai proses persiapan sebelum data dilakukan pemodelan. Pembersihan data teks perlu dilakukan untuk menghilangkan elemen yang tidak relevan atau tidak memberikan makna yang berarti. Tujuannya adalah agar proses pemodelan menjadi lebih efektif dan efisien. Pembersihan data mencakup proses tokenisasi, *replace slank* atau mengganti kata atau frasa yang bersifat slang atau bahasa gaul dengan kata atau frasa yang lebih baku, penghapusan emoji, *stemming*, *case folding*, *stopword*, penghapusan simbol-simbol, tanda baca dan unsur lainnya yang tidak terlalu memiliki kepentingan. Setelah dilakukan pembersihan data, maka data siap digunakan dalam proses pemodelan untuk memetakan data ke dalam topik-topik tertentu. Proses pembersihan data yang dilakukan dalam penelitian ini diilustrasikan seperti pada Tabel 1 dibawah ini.

Tabel 1. Tahap Pembersihan

Normalisasi	Sampel
Contoh	Semua seller amanah / sportif dlm mengirimkan barang
Tokenisasi	'Semua', 'seller', 'amanah', '/', 'sportif', 'dlm', 'mengirimkan', ' barang', ' '
Replace slank	Semua', 'penjual', 'amanah', '/', 'sportif', 'dalam', 'mengirimkan', ' barang', ' ' '
Penghapusan emoji	Semua', 'penjual', 'amanah', '/', 'sportif', 'dalam', 'mengirimkan', ' barang'
Stemming	Semua', 'penjual', 'amanah', '/', 'sportif', 'dalam', 'kirim', ' barang'
Case folding	semua', 'penjual', 'amanah', '/', 'sportif', 'dalam', 'kirim', ' barang'
Stopword	'penjual', 'amanah', '/', 'sportif', 'kirim', ' barang'
Simbol dan tanda baca	'penjual', 'amanah', 'sportif', 'kirim', ' barang'



Hasil akhir	penjual amanah sportif kirim barang
-------------	-------------------------------------

3.3 *Exploratory Data Analysis (EDA)*

Setelah melalui tahap pembersihan data, pada data teks ulasan aplikasi Shopee kemudian dilakukan analisis data eksploratif untuk mengeksplorasi struktur, pola, dan karakteristik dari data yang ada. Pada tahap ini, dilakukan beberapa analisis seperti analisis perolehan rating, frekuensi kata, dan visualisasi wordcloud untuk memperoleh *insight* atau pemahaman awal mengenai data yang ada.

3.4 *Pembobotan TF-IDF*

Term Frequency - Invers Document Frequency (TF-IDF) adalah suatu proses untuk memberikan bobot relasi suatu kata (*term*) dengan dokumen. Proses TF-IDF mengkombinasikan dua skema perhitungan bobot, yakni frekuensi kemunculan suatu kata di dalam suatu dokumen dan *inverse* frekuensi dokumen yang memuat kata tersebut. Metode TF-IDF menentukan frekuensi relatif kata-kata dalam dokumen tertentu melalui *invers* proporsi kata di seluruh korpus dokumen. *Invers Document Frequency* adalah metrik untuk menentukan seberapa jarang suatu kata didasarkan pada suatu dokumen [5]. Berikut adalah persamaan untuk menghitung nilai IDF dan bobot (w) masing-masing dokumen :

$$IDF_t = \log_{10} (D/df_i) \quad (1)$$

$$W_{d,t} = tf_{d,t} (IDF_t) \quad (2)$$

Keterangan :

D = jumlah dokumen yang berisi term (t)

df_i = jumlah kemunculan kata terhadap D

$W_{d,t}$ = bobot dokumen ke- d pada kata ke- t

$tf_{d,t}$ = frekuensi kata

3.5 *Pemodelan LDA dan Pengujian Parameter*

Latent Dirichlet Allocation (LDA) merupakan suatu model probabilistik yang digunakan untuk mengenali topik-topik yang terkandung dalam kumpulan dokumen teks. Model ini berdasarkan asumsi bahwa setiap dokumen terdiri dari kombinasi beberapa topik, dan setiap kata dalam dokumen berasal dari salah satu topik tersebut. LDA dapat digunakan untuk menyimpulkan, mengelompokkan, menghubungkan, dan memproses koleksi data besar karena menghasilkan daftar topik yang memiliki bobot untuk setiap dokumen dalam koleksi [6]. Dalam penelitian ini, parameter *passes* dan *iterations* dipertimbangkan untuk pemodelan dengan random state 142. Parameter *passes* mengontrol jumlah iterasi yang dilakukan oleh algoritma untuk melatih model LDA pada dataset yang diberikan. Sementara itu, *iterations* merujuk pada jumlah kali yang dilakukan oleh algoritma Gibbs *sampling* untuk menghasilkan perkiraan topik dari dokumen yang diberikan. Untuk menentukan nilai *passes* dan *iterations* final yang akan digunakan dalam penelitian ini, dilakukan pengujian dengan mengkombinasikan beberapa nilai *passes* dan *iterations*. Parameter *passes* dengan nilai 5, 10, 15, 20 diuji dengan kombinasi nilai *iterations* 50 dan 100.

3.6 *Pemilihan Model*

Jumlah topik (k) yang akan digunakan dalam model akhir dilihat dari nilai k -optimum yang paling sering muncul dari hasil kombinasi parameter. Setelah itu, parameter *passes* dan *iterations* yang memberikan rata-rata *coherence score* tertinggi dipilih sebagai model final (metrik *coherence score* yang digunakan adalah C_V). Dengan cara ini, model LDA yang dihasilkan diharapkan mampu memberikan representasi topik yang akurat dan optimal untuk dokumen yang diberikan.

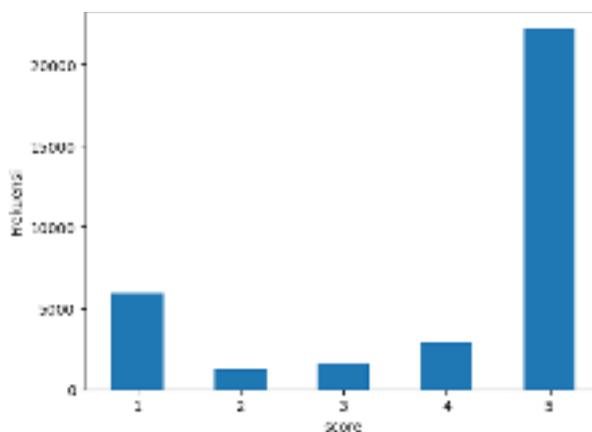
3.7 Interpretasi Topik dan Hasil

Setiap topik dalam LDA direpresentasikan sebagai distribusi kata, dan kata-kata yang memiliki probabilitas tertinggi dalam distribusi tersebut dianggap sebagai kata-kata yang paling merepresentasikan topik tersebut. Interpretasi topik dalam LDA dilakukan dengan melihat kata-kata yang paling berkontribusi dalam setiap topik. Setelah didapatkan topik-topik, kemudian model diaplikasikan pada keseluruhan data untuk menggali *insight* yang lebih mendalam.

III. HASIL DAN PEMBAHASAN

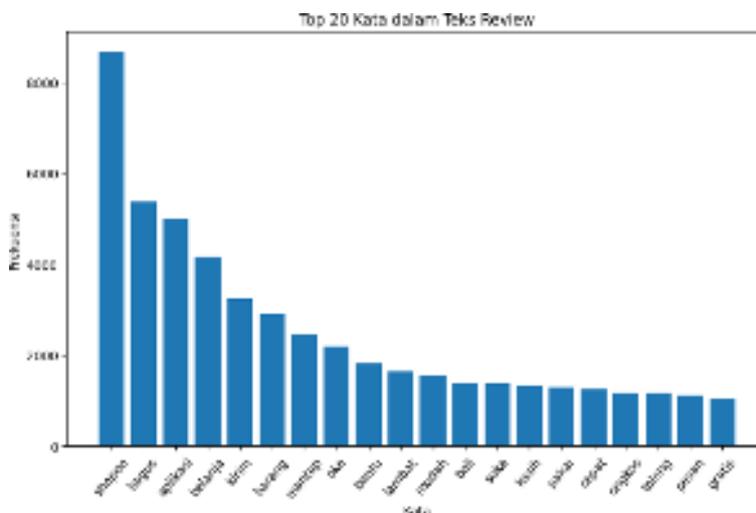
3.1. Data Ulasan Pengguna

Penelitian ini menggunakan data ulasan pengguna aplikasi Shopee di PlayStore yang terhitung sejak tanggal 1 April 2023 sampai 20 April 2023. Setelah dilakukan pembersihan pada data mentah, didapatkan 33.896 record data bersih yang akan masuk dalam proses pemodelan.



Gambar 2. Rating di PlayStore

Berdasarkan Gambar 2, terlihat bahwa distribusi rating yang pada aplikasi Shopee memiliki puncak tertinggi pada rating 5. Hal ini menunjukkan bahwa sebagian besar pengguna memberikan penilaian yang sangat baik terhadap Shopee. Namun, terdapat juga sejumlah pengguna yang memberikan rating 1 dan rating 2. Meskipun jumlah yang memberikan rating 1 dan rating 2 lebih sedikit dibandingkan dengan pengguna yang memberikan rating 5, perlu perhatian khusus untuk mengevaluasi alasan di balik rating tersebut dan memperbaiki masalah yang mungkin dihadapi oleh pengguna.



Gambar 3. Top 20 Kata dengan Frekuensi Tertinggi

Barplot pada Gambar 3 memberikan gambaran mengenai kata-kata yang sering muncul dalam keseluruhan data ulasan aplikasi Shopee di PlayStore yang dapat membantu dalam memahami hal yang diungkapkan pengguna secara umum. Kata ‘shopee’, ‘bagus’, ‘aplikasi’, ‘belanja’, ‘kirim’, merupakan lima kata teratas dengan kemunculan paling sering dalam data. Walaupun frekuensi kemunculan suatu kata dapat memberikan gambaran tentang seberapa sering kata tersebut digunakan dalam dokumen, tetapi ini tidak cukup digunakan untuk memahami topik yang sebenarnya dibahas oleh pengguna. Perlu dilakukan analisis lebih lanjut untuk memahami konteks kata-kata yang digunakan dalam dokumen ulasan tersebut.

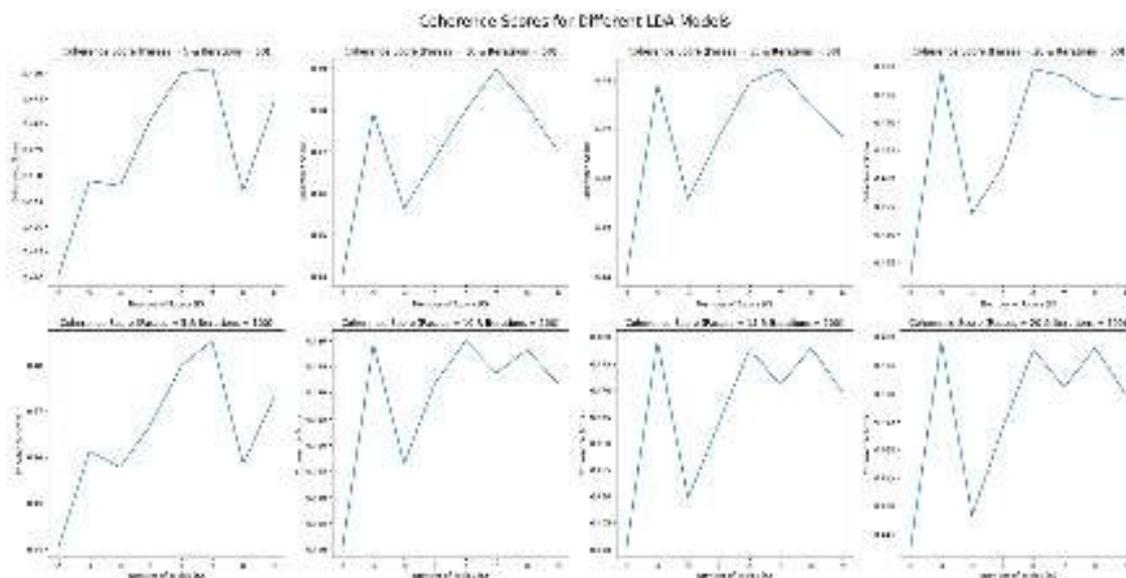


Gambar 4. Wordcloud

Visualisasi wordcloud pada Gambar 4 juga menampilkan beberapa kata kunci yang paling sering muncul dalam ulasan pengguna aplikasi Shopee. Dalam wordcloud tersebut juga terlihat hal-hal detail yang sebelumnya tidak diketahui, seperti kata ‘tolong’, ‘kecewa’, ‘lambat’.

4.2 Pemodelan

Penelitian ini mengimplementasikan algoritma *Latent Dirichlet Allocation* (LDA) pada data ulasan pengguna Shopee di PlayStore untuk melakukan topik modelling, yaitu mengidentifikasi topik-topik yang dibicarakan dalam ulasan pengguna dan menganalisis hubungan antar topik. Parameter *passes* dan *iterations* dipertimbangkan dalam pemodelan. Parameter *passes* dengan nilai 5, 10, 15, 20 diuji dengan kombinasi nilai *iterations* 50 dan 100 serta *random state* 142.



Gambar 5. Coherence Score untuk Setiap Parameter Pengujian

Jumlah topik (k) yang akan digunakan dalam model akhir dilihat dari nilai k -optimum yang paling sering muncul dari hasil kombinasi parameter. Setelah itu, parameter *passes* dan *iterations* yang memberikan rata-rata *coherence score* tertinggi dipilih sebagai parameter model final.

Tabel 2. Hasil Pengujian

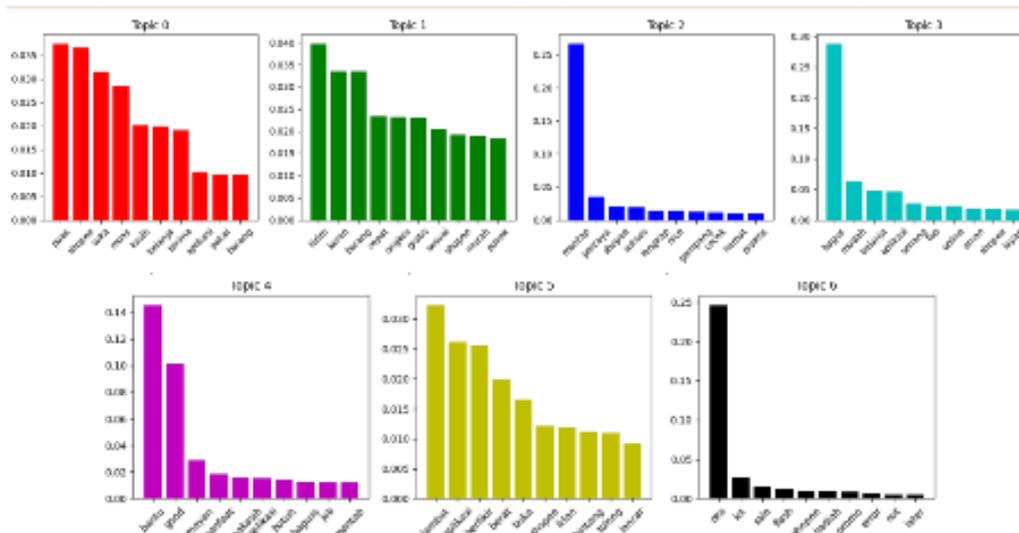
Passes	Iterations	Rata-rata	K-Optimum
5		5	
10	50	10	50
15		0.4686	7
20		0.468	6
5		0.4654	7
10	100	10	100
15		0.4661	3
20		0.4663	3

Tabel 3. Coherence Value (Passes = 10, Iterations = 50)

K	Coherence Value
2	0.4405
3	0.4793
4	0.4564
5	0.4684
6	0.4799
7	0.4898
8	0.4809
9	0.4703

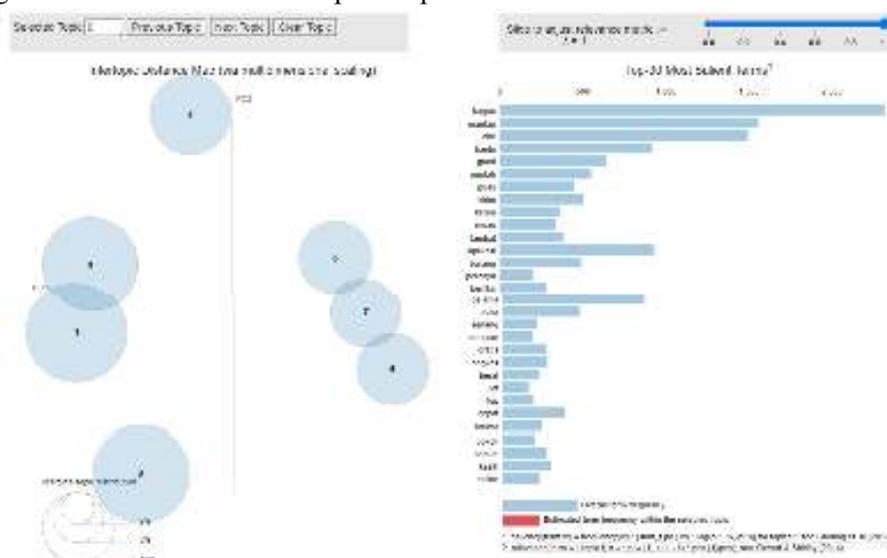
Hasil pengujian seperti yang ditunjukkan pada Tabel 7 menunjukkan bahwa nilai rata-rata *coherence score* tertinggi terdapat pada kombinasi nilai *passes* 10 dan nilai *iterations* 50 dengan nilai rata-rata *coherence score* sebesar 0.47068. *Coherence score* merupakan ukuran kualitas interpretasi model LDA. Oleh karena itu, memilih parameter model yang memberikan nilai rata-rata coherence

score tertinggi adalah langkah yang tepat menurut peneliti. Tabel 7 juga memberikan nilai K-Optimum yang paling sering muncul adalah 7. Oleh karena itu, nilai *passes* sebesar 10 dan nilai *iterations* sebesar 50 dipilih sebagai parameter model LDA final dengan jumlah topik sebanyak tujuh topik.



Gambar 6. Topik Utama berdasarkan Probabilitas Kata

Gambar 6 menunjukkan sebaran kata-kata di setiap topik yang terbentuk dan bobot kata-kata yang paling berkontribusi. Bobot ini menunjukkan seberapa penting kata tersebut dalam topik dan dihitung sebagai probabilitas kemunculan kata tersebut dalam topik. Sebagai contoh, untuk topik 0, kata-kata yang paling berkontribusi adalah “puas”, “shopee”, “suka”, “muas”, dan seterusnya, dengan bobot probabilitas tertentu untuk setiap kata. Semakin tinggi nilai probabilitasnya, semakin besar kemungkinan kata tersebut muncul pada topik tersebut.



Gambar 7. PyLDAvis secara Umum

Gambar 7 pada panel kiri menampilkan diagram gelembung dengan dimensi yang menggambarkan jarak antara topik dan area gelembung yang sesuai dengan prevalensi relatif topik di dalam korpus. Setiap gelembung mewakili satu topik, semakin besar gelembung maka semakin

tinggi persentase jumlah ulasan dalam korpus tentang topik tersebut. Jika melihat gelembung yang muncul pada tujuh topik yang terbentuk, dapat diketahui bahwa terdapat beberapa topik yang area gelembungnya memiliki ukuran lebih kecil dibandingkan topik lainnya. Oleh karena itu, frekuensi antara topik tersebut dengan topik lainnya sedikit berbeda dalam bobotnya. Selain itu juga terlihat beberapa gelembung yang saling beririsan, fenomena ini menunjukkan adanya kemungkinan bahwa antara topik yang satu dengan topik lainnya memiliki sedikit kesamaan atau kemiripan.

4.3 Interpretasi Topik

Berdasarkan hasil pemodelan yang telah dilakukan, ditemukan bahwa terdapat tujuh topik yang muncul pada data ulasan pengguna Shopee. Tiap topik memiliki sejumlah kata kunci yang memiliki bobot yang berbeda-beda. Bobot kata-kata tersebut menunjukkan tingkat kepentingan kata tersebut dalam konteks topik tertentu. Bobot ini dapat dianggap sebagai representasi numerik dari tingkat signifikansi kata dalam topik tersebut dibandingkan dengan kata-kata di topik lainnya. Bobot yang semakin tinggi menunjukkan semakin kuat hubungan antara kata-kata tersebut dengan topik yang sedang dibahas. Oleh karena itu, analisis bobot kata dalam topik dapat memberikan wawasan tentang topik yang dibicarakan dalam ulasan pengguna Shopee, dan dapat digunakan sebagai dasar untuk mengidentifikasi preferensi pelanggan dalam konteks pengalaman mereka dengan Shopee. Dengan mempertimbangkan bobot kemunculan kata di setiap topik dan diagram PyLDAvis, didapatkan inti dari setiap topik sebagai berikut :

- Topik pertama (0) terkait dengan **kepuasan pengguna**. Kata-kata seperti "puas", "suka", "kasih", dan "terima" menunjukkan kepuasan dari pengguna terhadap Shopee. Kepuasan pengguna merupakan faktor penting dalam bisnis *e-commerce*. Kepuasan pengguna dapat berdampak pada retensi pengguna, loyalitas, dan meningkatkan reputasi bisnis. Sehingga, topik ini perlu diperhatikan sebagai tolok ukur kepuasan pengguna.
- Topik kedua (1) dapat dikaitkan dengan aspek **pengiriman dan kesesuaian barang** yang diterima oleh pengguna. Terdapat kata-kata kunci seperti "kirim", "barang", dan "sesuai" yang menunjukkan bahwa para pengguna memberikan penilaian terhadap kualitas pengiriman dan kesesuaian barang yang mereka terima dari Shopee. Selain itu, kata-kata seperti "cepat", "ongkos", dan "gratis" juga muncul dalam topik ini, mengindikasikan bahwa para pengguna sangat memperhatikan waktu pengiriman dan biaya pengiriman yang dikenakan.
- Topik ketiga (2) menggambarkan **kepercayaan pengguna** terhadap Shopee. Bobot tertinggi pada topik ini adalah kata "mantap", "shopee", dan "percaya". Kata-kata ini merepresentasikan tingkat kepercayaan kepada Shopee. Kata-kata seperti "sukses", "nice", "gampang", dan "cocok" juga muncul pada topik ini, menunjukkan bahwa pengguna memandang Shopee sebagai platform *e-commerce* yang dapat diandalkan.
- Topik keempat (3) dapat diinterpretasikan sebagai topik yang berkaitan dengan **pengalaman pengguna** dalam berbelanja di Shopee. Bobot tertinggi pada topik ini didominasi oleh kata-kata seperti "bagus", "mudah", dan "belanja", yang menunjukkan kesan pengguna terhadap berbelanja di Shopee. Selain itu, kata-kata seperti "aman", "layan", dan "online" memberikan tambahan informasi bahwa topik ini berkaitan dengan pengalaman pengguna dalam berbelanja secara online di Shopee.
- Topik kelima (4) berkaitan dengan **layanan dan bantuan** yang diberikan oleh Shopee. Kata-kata yang paling sering muncul dalam topik ini adalah "bantu", "good", dan "lumayan". Bobot yang tinggi dari kata-kata tersebut menunjukkan bahwa pengguna memberikan penilaian terhadap layanan dan bantuan yang diberikan oleh Shopee. Disamping itu, terdapat kata-kata

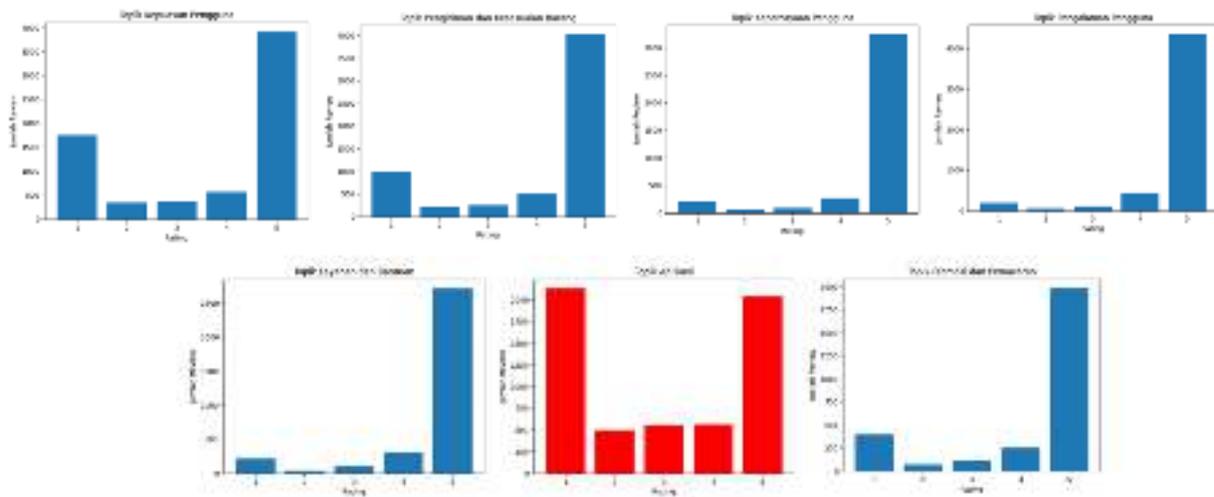
- seperti "manfaat", "terimakasih", dan "butuh" yang juga muncul dalam topik ini, menunjukkan bahwa pengguna merasa terbantu oleh layanan dan bantuan yang diberikan oleh Shopee.
- Topik keenam (5) berkaitan dengan **aplikasi Shopee**. Bobot tertinggi pada topik ini adalah kata "lambat" dan "aplikasi". Hal ini menunjukkan bahwa pengguna mengalami beberapa masalah dalam menggunakan aplikasi Shopee. Topik ini juga diperkuat dengan kata-kata “berat”, “iklan”, “tolong” yang mengindikasikan permasalahan saat menjalankan aplikasi.
 - Topik ketujuh (6) berkaitan dengan **promosi dan penawaran**. Berdasarkan bobot kata-kata yang ditemukan pada topik 6, dapat disimpulkan bahwa pengguna memberikan ulasan tentang promosi dan penawaran yang diberikan oleh Shopee. Kata-kata seperti "oke", "sale", "flash", "hadiah", dan "promo" menunjukkan bahwa pengguna memberikan feedback terkait promosi dan penawaran yang disediakan oleh Shopee.

Tabel 4. Kesimpulan Topik

Model	Kesimpulan Topik
Topic 0 : 0.037*"puas" + 0.037*"shopee" + 0.031*"suka" + 0.028*"muas" + 0.020*"kasih" + 0.020*"belanja" + 0.019*"terima" + 0.010*"aplikasi" + 0.010*"pakai" + 0.010*"barang"	Kepuasan Pengguna
Topic 1 : 0.040*"kirim" + 0.034*"keren" + 0.034*"barang" + 0.023*"cepat" + 0.023*"ongkos" + 0.023*"gratis" + 0.020*"sesuai" + 0.019*"shopee" + 0.019*"murah" + 0.018*"pokok"	Pengiriman dan Kesesuaian Barang
Topic 2 : 0.267*"mantap" + 0.035*"percaya" + 0.020*"shopee" + 0.020*"sukses" + 0.014*"lengkap" + 0.014*"nice" + 0.013*"gampang" + 0.011*"cocok" + 0.009*"hemat" + 0.009*"praktis"	Kepercayaan Pengguna
Topic 3 : 0.288*"bagus" + 0.063*"mudah" + 0.049*"belanja" + 0.047*"aplikasi" + 0.027*"senang" + 0.024*"top" + 0.023*"online" + 0.018*"aman" + 0.017*"shopee" + 0.017*"layan"	Pengalaman Pengguna
Topic 4 : 0.145*"bantu" + 0.102*"good" + 0.029*"lumayan" + 0.019*"manfaat" + 0.016*"terimakasih" + 0.015*"aplikasi" + 0.014*"butuh" + 0.013*"baguss" + 0.012*"job" + 0.012*"mantab"	Layanan dan Bantuan
Topic 5 : 0.032*"lambat" + 0.026*"aplikasi" + 0.025*"berfikir" + 0.020*"berat" + 0.017*"buka" + 0.012*"shopee" + 0.012*"iklan" + 0.011*"bintang" + 0.011*"tolong" + 0.009*"lancar"	Aplikasi
Topic 6 : 0.246*"oke" + 0.026*"lot" + 0.015*"sale" + 0.013*"flash" + 0.009*"shopee" + 0.009*"hadiah" + 0.008*"promo" + 0.007*"error" + 0.005*"out" + 0.005*"later"	Promosi dan Penawaran

4.4 Implementasi

Dalam tahap implementasi, model yang telah dibuat dan topik-topik yang terbentuk kemudian diterapkan pada keseluruhan data teks ulasan yang ada. Penerapan model pada data memberikan gambaran yang lebih komprehensif tentang penilaian pengguna terhadap setiap topik. Dari penerapan model pada data, dapat diketahui topik-topik mana saja yang perlu dijadikan perhatian khusus dilihat dari sebaran banyaknya penilaian yang kurang baik atau rating yang rendah yang diberikan pengguna.



Gambar 8. Sebaran Rating di Setiap Topik

Hasil dari implementasi model pada keseluruhan data memberikan pandangan yang lebih jelas dan mendalam mengenai topik-topik yang perlu dijadikan perhatian khusus. Dari Gambar 8 dapat disimpulkan bahwa topik aplikasi memiliki intensitas rating 1 paling tinggi, atau dalam pengertian lain pengguna kebanyakan memberikan rating 1 karena terkait dengan isu aplikasi. Kata "lambat", "aplikasi", "berat", "buka", merupakan kata-kata dengan bobot tinggi pada topik aplikasi. Hal ini menunjukkan bahwa aplikasi Shopee perlu diperbaiki dan ditingkatkan dalam hal kinerja sistem dan fungsionalitasnya. Selain itu, perlu diperhatikan bahwa topik kepuasan pengguna dan topik pengiriman dan kesesuaian barang memiliki persentase rating 1 yang juga cukup tinggi, Shopee perlu melakukan evaluasi dan perbaikan dalam meningkatkan tingkat kepuasan pengguna serta dalam hal pengiriman dan kesesuaian produk yang diterima pembeli.

IV. KESIMPULAN

Berdasarkan hasil dari proses analisis data dan pemodelan terhadap 33.896 ulasan pengguna aplikasi Shopee di PlayStore yang tercatat mulai dari tanggal 1 April 2023 sampai 21 April 2023 diperoleh kesimpulan sebagai berikut :

1. Distribusi rating yang pada aplikasi Shopee memiliki puncak tertinggi pada rating 5. Hal ini menunjukkan bahwa sebagian besar pengguna memberikan penilaian yang sangat baik terhadap Shopee. Kata "shopee", "bagus", "aplikasi", "belanja", "kirin", merupakan 5 kata teratas dengan kemunculan paling sering dalam data. Disamping itu, dalam wordcloud yang terbentuk juga terlihat hal-hal detail yang sebelumnya tidak tampak, seperti kata "tolong", "kecewa", "lambat" yang mengarah sesuatu yang negatif.
2. Pemodelan LDA dilakukan dengan mempertimbangkan parameter *passes* dan *iterations*. Parameter *passes* dengan nilai 5, 10, 15, 20 diuji dengan kombinasi nilai *iterations* 50 dan 100 serta *random state* 142. Hasil pengujian menunjukkan bahwa nilai rata-rata coherence score tertinggi terdapat pada kombinasi nilai *passes* 10 dan nilai *iterations* 50 dengan nilai rata-rata *coherence score* sebesar 0.47068. Nilai K-Optimum yang paling sering muncul dari pengujian tersebut adalah 7. Oleh karena itu, nilai *passes* sebesar 10 dan nilai *iterations* sebesar 50 dipilih sebagai parameter model LDA final dengan jumlah topik sebanyak tujuh topik. Adapun topik-topik yang terbantu antara lain topik kepuasan pengguna, topik pengiriman dan kesesuaian



- barang, topik kepercayaan pengguna, topik pengalaman pengguna, topik layanan dan bantuan, topik aplikasi, dan topik terkait promosi dan penawaran.
3. Hasil dari implementasi model pada keseluruhan data, terlihat bahwa topik "aplikasi" menjadi topik yang paling kritis dan mendapatkan rating 1 terbanyak. Lebih lanjut, kata-kata dengan bobot tinggi pada topik aplikasi seperti "lambat", "berat", dan "buka" menunjukkan bahwa masalah terkait kinerja sistem dan fungsionalitas aplikasi menjadi salah satu isu yang perlu diperbaiki dan ditingkatkan.

UCAPAN TERIMA KASIH

REFERENSI

1. R. Ameliah, R. Adi Hegara, I. Rahmawati, dan Dkk, “Status Literasi Digital di Indonesia Ringkasan Eksekutif,” *Indeks Literasi Digit. Indones.*, hal. 9, 2021, [Daring]. Tersedia pada: <https://katadata.co.id/StatusLiterasiDigital>.
2. M. Habibi, A. Priadana, A. B. Saputra, dan P. W. Cahyo, “Topic Modelling of Germas Related Content on Instagram Using Latent Dirichlet Allocation (LDA),” vol. 34, no. Ahms 2020, hal. 260–264, 2021, doi: 10.2991/ahsr.k.210127.060.
3. A. R. Destarani, I. Slamet, dan S. Subanti, “Trend Topic Analysis using Latent Dirichlet Allocation (LDA) (Study Case: Denpasar People’s Complaints Online Website),” *J. Ilm. Tek. Elektro Komput. dan Inform.*, vol. 5, no. 1, hal. 50–58, 2019, doi: 10.26555/jiteki.v5i1.13088.
4. I. Sutherland, Y. Sim, S. K. Lee, J. Byun, dan K. Kiatkawsin, “Topic modeling of online accommodation reviews via latent dirichlet allocation,” *Sustain.*, vol. 12, no. 5, hal. 1–15, 2020, doi: 10.3390/su12051821.
5. A. S. Neogi, K. A. Garg, R. K. Mishra, dan Y. K. Dwivedi, “Sentiment analysis and classification of Indian farmers’ protest using twitter data,” *Int. J. Inf. Manag. Data Insights*, vol. 1, no. 2, hal. 100019, 2021, doi: 10.1016/j.jjime.2021.100019.
6. J. C. Campbell, A. Hindle, dan E. Stroulia, “Latent Dirichlet Allocation: Extracting Topics from Software Engineering Data,” *Art Sci. Anal. Softw. Data*, vol. 3, hal. 139–159, 2015, doi: 10.1016/B978-0-12-411519-4.00006-9.
7. D. Endriani, “Analisis topic modelling mengenai pemberlakuan pembatasan kegiatan masyarakat menggunakan latent direchlet allocation (LDA),” *Univ. Islam Indonesia.*, hal. 1–79, 2022.