

PEMODELAN TOPIK DAN ANALISIS SENTIMEN ULASAN PENGGUNA APLIKASI MOBILE JKN MENGGUNAKAN ALGORITMA LATENT DIRICHLET ALLOCATION DAN RANDOM FOREST

Anis Raudhatul Eni, Dian Christien Arisona, Gusti Ngurah Adhi Wibawa,
Agusrawati, Irma Yahya, Bahriddin Abapihi
Statistika, Universitas Halu Oleo
Jln. H.E.A Mokodompit Kendari, Indonesia
anisraudhatuleni@gmail.com

ABSTRAK

Aplikasi Mobile JKN merupakan aplikasi layanan digital yang dikembangkan oleh BPJS Kesehatan untuk mendukung pelaksanaan program Jaminan Kesehatan Nasional (JKN) di Indonesia. Sebagai aplikasi yang digunakan luas oleh masyarakat, ulasan pengguna di Google Play Store menjadi sumber informasi penting yang perlu dianalisis untuk perbaikan layanan. Penelitian ini bertujuan untuk menganalisis sentimen pengguna dan mengidentifikasi topik utama dalam ulasan aplikasi Mobile JKN. Sebanyak 9.677 ulasan yang telah dipraolah dianalisis menggunakan metode *Latent Dirichlet Allocation* (LDA) untuk pemodelan topik dan algoritma *Random Forest* untuk klasifikasi sentimen. Hasil analisis sentimen menunjukkan bahwa 11,1% ulasan bersentimen positif, sementara 88,9% lainnya bersentimen negatif. LDA mengidentifikasi delapan topik utama dengan nilai *coherence score* sebesar 0,575. Topik seperti OTP, verifikasi wajah, dan konektivitas didominasi oleh sentimen negatif, sementara fitur administrasi relatif lebih positif. Model *Random Forest* mencapai akurasi sebesar 96% dengan performa sangat baik untuk sentimen negatif, yaitu presisi 97%, *recall* 99%, dan *F1-Score* sebesar 98%. Untuk sentimen positif, model menunjukkan presisi sebesar 86%, *recall* 72%, dan *F1-Score* sebesar 78%.

Kata kunci : Mobile JKN, Analisis Sentimen, Pemodelan Topik, Random Forest, LDA.

1. PENDAHULUAN

Kesehatan adalah salah satu hal mendasar yang dibutuhkan setiap manusia. Kondisi yang sehat memungkinkan manusia untuk beraktivitas dengan maksimal. Ini menunjukkan bahwa kesehatan juga menjadi pendorong kemajuan bangsa. Salah satu bentuk tanggung jawab pemerintah untuk memenuhi hak hidup sehat bagi masyarakat adalah dengan membuat program Jaminan Kesehatan Nasional (JKN). JKN diselenggarakan oleh Badan Penyelenggara Jaminan Sosial (BPJS) Kesehatan. Pemerintah Indonesia melaksanakan program JKN sebagai bentuk komitmen untuk mencapai tujuan pembangunan berkelanjutan (SDGs) dengan memastikan tercapainya cakupan kesehatan semesta [1].

Program JKN mulai dilaksanakan sejak 1 Januari 2014. BPJS Kesehatan kemudian terus melakukan inovasi terkait program ini, salah satunya dengan meluncurkan aplikasi Mobile JKN pada tanggal 15 September 2017. Mobile JKN adalah aplikasi *mobile* yang dikembangkan untuk memudahkan peserta JKN dalam mengakses berbagai layanan kesehatan. Aplikasi ini dilengkapi berbagai fitur yang diharapkan dapat memudahkan dan mengefisiensikan waktu peserta JKN. Aplikasi ini menjadi sangat penting bagi sistem pelayanan kesehatan khususnya terhadap peserta JKN karena kebijakan beberapa fasilitas kesehatan yang mewajibkan peserta JKN untuk menggunakan aplikasi ini. Berdasarkan informasi yang dilansir dari *website* resmi RSUD Ir. Soekarno Kabupaten Sukoharjo, peserta JKN yang ingin

melakukan rawat jalan wajib melakukan pendaftaran *online* melalui aplikasi Mobile JKN [2].

Mobile JKN sudah tersedia dan dapat diunduh pada Google Play Store. Aplikasi ini telah diunduh lebih dari 50 juta kali di Google Play Store dan mendapatkan rating 4,3 dari 5, serta 750.000 ulasan pengguna. Ulasan ini dapat memberikan informasi mengenai topik-topik pembahasan dan sentimen penggunaannya. Pemodelan topik dan analisis sentimen dapat dilakukan untuk menemukan topik serta sentimen dalam ulasan [3]. Informasi topik dan sentimen yang terkandung dalam ulasan dapat membantu dalam meningkatkan kepuasan dan memperbaiki layanan [4].

Pemodelan topik adalah model statistik yang digunakan untuk menemukan topik abstrak dalam kumpulan dokumen. Salah satu metode dalam pemodelan topik adalah *Latent Dirichlet Allocation* (LDA). LDA adalah metode statistika yang digunakan untuk menganalisis topik dalam dokumen dengan cara merepresentasikan dalam bentuk probabilitas tertentu [5].

Analisis sentimen adalah metode yang digunakan untuk mengungkapkan opini dan sentimen dalam suatu teks [4]. Analisis ini bertujuan untuk melihat opini masyarakat terhadap suatu hal. Analisis sentimen dapat dilakukan dengan bantuan *machine learning*, salah satunya adalah menggunakan algoritma *Random Forest*. Metode *Random Forest* adalah sebuah metode yang menggunakan pohon keputusan sebagai dasar klasifikasi yang dibangun dan dikombinasikan. Kelebihan metode *Random Forest* dibanding metode

lainnya adalah hasil akurasi yang baik, relatif kuat terhadap *outliers* dan *noise*, lebih cepat dibandingkan dengan *bagging* dan *boosting*, dan sifatnya sederhana dan mudah diparalelkan [6].

Penelitian ini bertujuan untuk menggambarkan sentimen dan topik yang dibahas oleh pengguna aplikasi Mobile JKN, serta mengidentifikasi proporsi sentimen pada setiap topik. Untuk mencapai tujuan tersebut, penelitian ini menerapkan metode *Latent Dirichlet Allocation* (LDA) guna mengungkap topik-topik utama dalam ulasan pengguna, dan algoritma *Random Forest* untuk mengklasifikasikan sentimen yang terkandung dalam masing-masing ulasan.

2. TINJAUAN PUSTAKA

Pada tahapan ini, akan dipaparkan beberapa literatur ilmiah yang relevan dengan penelitian.

2.1. Penelitian Terdahulu

Penelitian [7] mengenai pemodelan topik pada media berita daring dengan studi kasus merek Somethinc berhasil mengidentifikasi enam topik menggunakan metode LDA, dengan nilai *coherence score* optimal sebesar 0,404. Masing-masing topik direpresentasikan oleh daftar probabilitas kata yang paling mewakili topik tersebut, yang selanjutnya diinterpretasikan berdasarkan penilaian manusia.

Penelitian [4] menunjukkan bahwa akurasi klasifikasi dengan menggunakan *Random Forest* lebih tinggi jika dibandingkan dengan akurasi *Support Vector Machine*, yaitu sebesar 100% jika menggunakan *Random Forest* dan 99% ketika menggunakan *Support Vector Machine*.

Penelitian [3] yang memadukan pemodelan topik dan analisis sentimen pada aplikasi EdLink berhasil menemukan 3 aspek tersembunyi yang dibahas dalam ulasan aplikasi, yaitu *usability*, *reliability*, dan *performance efficiency*. Ketiga aspek ini didominasi oleh ulasan bersentimen negatif, dengan peringkat tertinggi terdapat pada aspek *reliability*, selanjutnya *performance efficiency*, dan terakhir *usability*.

2.2. Mobile JKN

Mobile JKN adalah aplikasi digital yang dibuat oleh BPJS Kesehatan dan diluncurkan pada tanggal 15 September 2017. Aplikasi ini adalah bentuk inovasi dari BPJS Kesehatan untuk memudahkan peserta JKN dalam mengakses layanan Jaminan Kesehatan Nasional-Kartu Indonesia Sehat (JKN-KIS). Aplikasi ini juga ditujukan untuk memudahkan pengguna dalam mengakses layanan JKN-KIS [8].

2.3. Analisis Sentimen

Analisis sentimen adalah riset komputasional dari opini, sentimen, dan emosi yang diekspresikan secara tekstual [9]. Analisis sentimen pada dasarnya bekerja dengan mengukur polaritas suatu teks. Dalam analisis sentimen, suatu teks dapat memiliki polaritas positif atau negatif [10]. Hasil analisis sentimen dapat digunakan sebagai informasi kualitas suatu layanan

[11]. Dalam konteks ulasan aplikasi, sentimen positif dapat memberikan informasi bahwa aplikasi memiliki kualitas yang baik dan memuaskan. Sebaliknya, sentimen negatif dapat memberikan informasi bahwa aplikasi dinilai kurang baik, tidak memuaskan, dan mengalami masalah yang perlu diperbaiki.

2.4. Latent Dirichlet Allocation (LDA)

Latent Dirichlet Allocation (LDA) adalah model statistik yang digunakan untuk menemukan topik-topik tersembunyi dalam kumpulan dokumen. LDA berusaha untuk melihat dokumen dengan cara mundur untuk menemukan satu set topik yang mungkin telah dikoleksi. LDA menganggap suatu dokumen direpresentasikan dengan berbagai topik dengan probabilitas tertentu [5]. Penentuan setiap kata yang masuk ke dalam topik tertentu dilakukan melalui proses distribusi probabilitas yang memanfaatkan Gibbs Sampling [12].

LDA mengasumsikan jumlah topik dalam suatu dokumen telah diketahui. Jika tidak ada informasi terkait hal ini maka penentuan jumlah topik dapat dilakukan dengan melihat *coherence score*. *Coherence score* adalah ukuran untuk menentukan jumlah topik dengan nilai tertinggi yang akan diambil [13].

2.5. Random Forest

Random Forest adalah metode klasifikasi yang menggunakan beberapa pohon keputusan dan penentuan keputusan klasifikasinya didasarkan pada suara terbanyak dari pohon-pohon tersebut [11].

Pohon keputusan terdiri dari *root node*, *internal node*, dan *leaf node*. *Root node* merupakan simpul yang terletak paling atas. *Internal node* adalah simpul percabangan, dimana *node* ini mempunyai *output* minimal dua dan hanya ada satu input. Sedangkan *leaf node* merupakan simpul terakhir yang hanya memiliki satu input dan tidak mempunyai *output*.

Penentuan proses akhir klasifikasi dengan metode *Random Forest* dapat menggunakan nilai ketidakmurnian gini tertimbang untuk menentukan variabel *split* yang akan dijadikan *root* atau *node*. Atribut dengan nilai ketidakmurnian gini tertimbang paling kecil akan dipilih menjadi *split*. Perhitungan ketidakmurnian gini dan ketidakmurnian gini tertimbang terhadap klasifikasi dilakukan menggunakan persamaan berikut:

$$Gini = 1 - \sum_{i=1}^C \pi_i^2 \quad (1)$$

$$Gini_{split} = \sum_{j=1}^K \left(\frac{n_j}{n} \times Gini_j \right) \quad (2)$$

Keterangan:

C : Jumlah kelas dalam suatu node

π_i : Probabilitas relatif kelas ke- i dalam node tersebut

K : Jumlah *node* hasil pemisahan

n_j : Jumlah data pada *node* ke- j
 n : Jumlah total data pada *parent node*
 $Gini_j$: nilai gini indeks pada *node* ke- j

2.6. Evaluasi Model

Evaluasi model memiliki peran yang sangat penting karena digunakan untuk mengukur kinerja dari hasil kerja algoritma yang digunakan. Salah satu bentuk evaluasi model dalam *machine learning* adalah matriks konfusi. Matriks konfusi mengandung informasi berupa perbandingan hasil klasifikasi yang dilakukan oleh sistem dengan hasil klasifikasi sebenarnya. Matriks konfusi dilakukan untuk mengetahui sejauh mana model bekerja sesuai dengan apa yang diinginkan [11]. Bentuk matriks konfusi dapat dilihat dalam Tabel 1 berikut:

Tabel 1. Matriks Konfusi

Prediksi	Aktual	
	TRUE	FALSE
TRUE	True Positive (TP)	False Negative (FN)
FALSE	True Negative (TN)	False Positive (FP)

Nilai-nilai pada matriks konfusi dapat digunakan untuk menghitung nilai akurasi, presisi, dan *recall*. Akurasi merupakan rasio prediksi benar (positif dan negatif) dengan keseluruhan data. Persamaan akurasi dapat dilihat sebagai berikut:

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

Presisi adalah rasio prediksi benar positif dibandingkan dengan keseluruhan hasil yang diprediksi positif. Persamaannya dapat ditulis sebagai berikut:

$$Presisi = \frac{TP}{(TP + FP)} \quad (4)$$

Recall adalah rasio prediksi benar positif dibandingkan dengan keseluruhan data aktual positif. Persamaan *recall* adalah sebagai berikut:

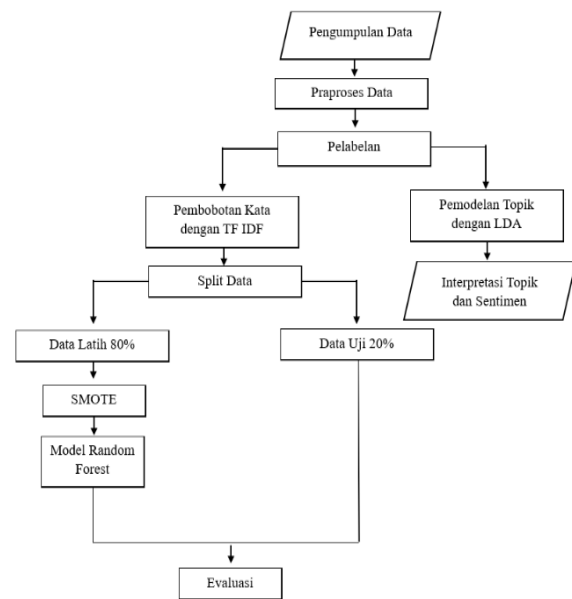
$$Recall = \frac{TP}{(TP + FN)} \quad (5)$$

F1-Score merupakan metrik evaluasi yang mencerminkan keseimbangan antara presisi dan *recall*. Persamaan *F1-Score* adalah sebagai berikut:

$$F1 - Score = 2 \times \frac{presisi \times recall}{presisi + recall} \quad (6)$$

3. METODE PENELITIAN

Pada bagian ini memuat metode yang digunakan pada pembuatan skripsi.



Gambar 1. Diagram Alir Penelitian

3.1. Pengumpulan Data

Data pada penelitian ini merupakan data ulasan aplikasi Mobile JKN pada Google Play Store yang diperoleh dengan cara *scrapping* menggunakan *API Google-Play-Scraper*. Variabel yang digunakan adalah variabel *username* dan *content* yang merupakan nama pengguna dan isi ulasan.

3.2. Praproses Data

Praproses data menjadi tahap awal dalam *text mining* untuk mempersiapkan data teks sebelum digunakan pada proses selanjutnya. Praproses data pada penelitian ini meliputi data *cleansing*, *case folding*, tokenisasi, *stopword removal*, normalisasi, dan *stemming*.

3.3. Pelabelan Sentimen

Pelabelan sentimen adalah proses memberi label dan mengelompokkan ulasan ke dalam kelas sentimen negatif dan positif. Pada penelitian ini, pelabelan sentimen dilakukan menggunakan metode Inset Leksikon, yaitu leksikon yang dibuat dengan analisis data *twitter*. Beberapa modifikasi dilakukan terhadap leksikon tersebut, salah satunya adalah penghapusan kata-kata yang muncul pada kedua kategori (positif dan negatif), dengan mempertimbangkan kesesuaian makna dalam konteks ulasan aplikasi Mobile JKN.

3.4. Pemodelan Topik

Pemodelan topik pada penelitian ini menggunakan metode LDA. LDA dilakukan dengan bantuan *package tomotopy* pada *Google Colab*. LDA akan menghasilkan distribusi probabilitas topik pada tiap dokumen, dan distribusi probabilitas kata pada tiap topik.

3.5. Klasifikasi Sentimen menggunakan Random Forest

Klasifikasi sentimen dengan *Random Forest* merupakan proses pemodelan untuk memprediksi label sentimen berdasarkan fitur yang diekstraksi dari teks ulasan. Proses klasifikasi sentimen dalam penelitian ini terdiri dari beberapa tahapan, yaitu ekstraksi fitur dengan TF-IDF, membagi data menjadi 80% data latih dan 20% data uji, menerapkan teknik SMOTE pada data latih untuk menangani ketidakseimbangan kelas, serta mengevaluasi performa model menggunakan data uji.

3.6. Interpretasi Topik dan Sentimen

Metode LDA menghasilkan distribusi probabilitas dari sejumlah topik yang muncul pada setiap ulasan. Ulasan-ulasan tersebut kemudian dikategorikan ke dalam topik yang memiliki probabilitas kemunculan tertinggi. Setelah itu, ulasan dalam masing-masing topik dianalisis lebih lanjut berdasarkan sentimennya, yaitu apakah bersifat negatif atau positif.

4. HASIL DAN PEMBAHASAN

Memuat hasil, Pengujian dan pembahasan tentang skripsi yang telah dilakukan.

4.1. Pengumpulan Data

Pengumpulan data pada penelitian ini menggunakan teknik *web scraping*. Pengumpulan data ulasan aplikasi pada Google Play Store dengan teknik *web scraping* dapat dilakukan dengan bantuan *API Google-Play-Scraper*. 9868 data ulasan pengguna Mobile JKN pada Google Play Store berhasil dikumpulkan pada tanggal 15 Januari 2025.

	userName	at	content
1			
2	Pengguna 2025-01-1		Sudah beberapa kali mau daftar via jkn tp gak bisa2 Krn no otp TDK dikirim2. Tolong diperbaiki lagi aplikasinya. Tidak menyusahkan orang yg lagi sakit. Sampe skrg saya gak bisa daftar.
3	Pengguna 2019-09-0		aplikasi sangat membantu sekali, pengaduan pelayanan juga baik dan direspon cepat. tapi
4	Pengguna 2025-01-0		Menurut saya aplikasi mempersulit proses pendaftaran sangat merepotkan saya tidak bisa
5	Pengguna 2025-01-0		Capek banget dari tadi mau mendaftar aja sulit banget. Ngebug terus Susah banget buat
6	Pengguna 2025-01-1		Perlu di tingkatkan lagi untuk performa apk, karena semua pengguna tidak sepenuhnya me
7	Pengguna 2025-01-1		Aplikasi yg tidak dapat di gunakan, tidak ada fungsi. Hanya untuk verifikasi wajah saja susa
8	Pengguna 2024-12-2		Setelah saya memberikan penilaian kurang baik ke dokter nya yang menurut saya kurang
9	Pengguna 2024-12-2		tolong perbaiki masalah di daftar masalah kyc nya kasihan yang mau daftar, masih banyak

Gambar 2. Data ulasan pengguna Mobile JKN

Gambar 2 menunjukkan data ulasan pengguna aplikasi Mobile JKN yang diperoleh dari hasil *scraping* dengan pengaturan *most relevant*. Data disimpan dalam format *xlsx* dengan total data sebanyak 9868 data.

4.2. Praproses Data

Praproses data dilakukan dengan bantuan *library* yang disediakan oleh Google Colab dan melalui beberapa tahapan meliputi data *cleansing*, *case folding*, tokenisasi, normalisasi, *stopword removal*, *stemming* dan *filtering*.

4.2.1. Data Cleansing

Tahapan ini bertujuan membersihkan data dari serangkaian elemen yang tidak relevan seperti tagar, simbol atau karakter, tanda baca, dan emotikon pada kalimat, serta memastikan tidak ada duplikat dan data kosong. Dari data ternyata terdapat data duplikat yang dapat dilihat pada Tabel 2.

Tabel 2. Data Cleansing (Duplikat)

Waktu Ulasan Diunggah	Sebelum	Sesudah
10/01/2025 03:25	Verifikasi wajah tidak bisa	Verifikasi wajah tidak bisa
01/12/2024 11:05	Verifikasi wajah tidak bisa	

Hasil pembersihan data dari serangkaian elemen yang tidak relevan seperti tagar, simbol atau karakter, tanda baca, dan emotikon dapat dilihat pada Tabel 3.

Tabel 3. Data Cleansing

Sebelum	Sesudah
Sudah beberapa kali mau daftar via jkn tp gak bisa2 Krn no otp TDK dikirim2. Tolong diperbaiki lagi aplikasinya. Tidak menyusahkan orang yg lagi sakit. Sampe skrg saya gak bisa daftar.	Sudah beberapa kali mau daftar via jkn tp gak Krn no otp TDK Tolong diperbaiki lagi aplikasinya Tidak menyusahkan orang yg lagi sakit Sampe skrg saya gak bisa daftar

4.2.2. Case Folding

Case folding adalah tahapan yang dilakukan untuk menyeragamkan seluruh huruf dalam ulasan ke dalam bentuk huruf kecil. Hasil dari proses *case folding* dapat dilihat pada Tabel 4.

Tabel 4. Case Folding

Sebelum	Sesudah
Sudah beberapa kali mau daftar via jkn tp gak Krn no otp TDK Tolong diperbaiki lagi aplikasinya Tidak menyusahkan orang yg lagi sakit Sampe skrg saya gak bisa daftar	sudah beberapa kali mau daftar via jkn tp gak krn no otp tdk tolong diperbaiki lagi aplikasinya tidak menyusahkan orang yg lagi sakit sampe skrg saya gak bisa daftar

4.2.3. Tokenisasi

Tokenisasi adalah tahapan pemecahan kalimat ke dalam bentuk kata per kata. Hasil dari proses tokenisasi dapat dilihat pada Tabel 5.

Tabel 5. Tokenisasi

Sebelum	Sesudah
Sudah beberapa kali mau daftar via jkn tp gak Krn no otp TDK Tolong diperbaiki lagi aplikasinya Tidak menyusahkan orang yg lagi sakit Sampe skrg saya gak bisa daftar	sudah beberapa kali mau daftar via jkn tp gak krn no otp tdk tolong diperbaiki lagi aplikasinya tidak menyusahkan orang yg lagi sakit sampe skrg saya gak bisa daftar

4.2.4. Normalisasi

Normalisasi adalah proses mengubah kata menjadi bentuk standar. Hasil dari proses normalisasi dapat dilihat pada Tabel 6.

Tabel 6. Normalisasi

Sebelum	Sesudah
['sudah', 'beberapa', 'kali', 'mau', 'daftar', 'via', 'jkn', 'tapi', 'gak', 'krn', 'no', 'otp', 'tidak', 'tolong', 'diperbaiki', 'lagi', 'aplikasinya', 'tidak', 'menyusahkan', 'orang', 'yg', 'lagi', 'sakit', 'sampe', 'skrg', 'saya', 'gak', 'bisa', 'daftar']	['sudah', 'beberapa', 'kali', 'mau', 'daftar', 'via', 'jkn', 'tapi', 'tidak', 'karena', 'nomor', 'otp', 'tidak', 'tolong', 'diperbaiki', 'lagi', 'aplikasinya', 'tidak', 'menyusahkan', 'orang', 'yang', 'lagi', 'sakit', 'sampai', 'sekarang', 'saya', 'tidak', 'bisa', 'daftar']

4.2.5. Stopword Removal

Stopword removal berfungsi untuk menghapus kata-kata yang tidak memiliki makna penting dalam ulasan. Hasil dari proses *stopword removal* dapat dilihat pada Tabel 7.

Tabel 7. Stopword Removal

Sebelum	Sesudah
['sudah', 'beberapa', 'kali', 'mau', 'daftar', 'via', 'jkn', 'tapi', 'tidak', 'karena', 'nomor', 'otp', 'tidak', 'tolong', 'diperbaiki', 'lagi', 'aplikasinya', 'tidak', 'menyusahkan', 'orang', 'yang', 'lagi', 'sakit', 'sampai', 'sekarang', 'saya', 'tidak', 'bisa', 'daftar']	['kali', 'daftar', 'via', 'jkn', 'tidak', 'nomor', 'otp', 'tidak', 'tolong', 'diperbaiki', 'aplikasinya', 'tidak', 'menyusahkan', 'orang', 'sakit', 'tidak', 'daftar']

4.2.6. Stemming dan Filtering

Stemming adalah tahapan mengubah semua kata ke dalam bentuk kata dasarnya. Tahapan ini berperan penting untuk menyederhanakan analisis selanjutnya dengan cara menghilangkan imbuhan yang tidak diperlukan. Hasil dari proses *stemming* dan *filtering* dapat dilihat pada Tabel 8.

Tabel 8. Stemming dan Filtering

Sebelum	Sesudah
['kali', 'daftar', 'via', 'jkn', 'tidak', 'nomor', 'otp', 'tidak', 'tolong', 'diperbaiki', 'aplikasinya', 'tidak', 'menyusahkan', 'orang', 'sakit', 'tidak', 'daftar']	kali daftar via jkn tidak nomor otp tidak tolong diperbaiki aplikasi tidak susah orang sakit tidak daftar

4.3. Pelabelan Sentimen

Pelabelan dilakukan dengan menggunakan Inset Leksikon yang dibuat oleh Fajri Koto dan Rahmaningtyas. Hasil pelabelan menunjukkan terdapat 8600 ulasan bersentimen negatif, 1077 ulasan bersentimen positif, dan 191 ulasan bersentimen netral. Hasil pelabelan dapat dilihat pada Tabel 9.

Tabel 9. Hasil Pelabelan

Label	Jumlah
Negatif	8600
Positif	1077
Netral	191

Mengingat penelitian ini berfokus untuk melihat keluhan dan pujian masyarakat terhadap aplikasi Mobile JKN maka ulasan yang berlabel netral akan dihapus dari dataset. Hasil pelabelan setelah penghapusan data dengan label netral dapat dilihat pada Tabel 10.

Tabel 10. Hasil Pelabelan setelah Penghapusan Netral

Label	Jumlah
Negatif	8600
Positif	1077

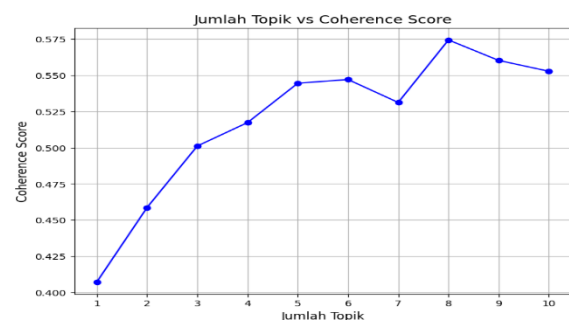
4.4. Pemodelan Topik

Dalam penelitian ini, dilakukan pemodelan topik dengan LDA untuk menemukan topik-topik yang terdapat dalam ulasan pengguna Mobile JKN. Topik-topik ini kemudian dapat memberi informasi mengenai fitur, layanan, pengalaman, keluhan yang dirasakan oleh pengguna aplikasi Mobile JKN.

4.4.1. Penentuan Jumlah Topik

LDA mengasumsikan jumlah topik telah diketahui sebelumnya, sehingga perlu dideklarasikan di awal. Jika tidak memiliki informasi mengenai jumlah topik yang terdapat dalam kumpulan dokumen, maka dapat dilakukan pemilihan jumlah topik optimal menggunakan nilai *coherence score*. Jumlah topik dengan *coherence score* tertinggi akan dipilih menjadi jumlah topik dalam pemodelan pada penelitian ini.

Dilakukan *hyperparameter tuning* dengan mencoba berbagai kombinasi jumlah topik, yaitu 1 hingga 10 untuk melihat jumlah topik mana yang memiliki *coherence score* tertinggi.



Gambar 3. Grafik Coherence Score vs Jumlah Topik

Berdasarkan Gambar 3, diketahui jumlah topik 8 menghasilkan *coherence score* tertinggi yaitu sebesar 0,575. Oleh karena itu, jumlah topik 8 dipilih dalam penelitian ini.

4.4.2. Hasil Pemodelan Topik

Setelah menentukan jumlah topik maka dilakukan pemodelan topik berdasarkan jumlah topik

yang telah ditentukan. Pemodelan topik ini menghasilkan distribusi kata untuk tiap topik. Hasil

distribusi kata untuk tiap topik dapat dilihat pada Tabel 11.

Tabel 11. Distribusi Kata tiap Topik

Topik	Kata
0	0.141*"tidak" + 0.051*"root" + 0.051*"buka" + 0.043*"jkn" + 0.035*"jalan" + 0.032*"emulator" + 0.032*"mobile" + 0.026*"muncul" + 0.023*"update" + 0.022*"perangkat"
1	0.071*"tidak" + 0.038*"update" + 0.031*"tolong" + 0.024*"buka" + 0.023*"nya" + 0.023*"coba" + 0.022*"aplikasi" + 0.016*"diperbaiki" + 0.015*"proses" + 0.014*"perbaiki"
2	0.048*"mudah" + 0.034*"jkn" + 0.031*"sehat" + 0.030*"tidak" + 0.029*"bantu" + 0.028*"layan" + 0.027*"bpjs" + 0.022*"fasilitas" + 0.021*"data" + 0.021*"ubah"
3	0.085*"daftar" + 0.074*"tidak" + 0.036*"bayar" + 0.031*"antri" + 0.025*"online" + 0.022*"rujuk" + 0.018*"nya" + 0.016*"ambil" + 0.016*"kontrol" + 0.015*"sehat"
4	0.087*"nomor" + 0.084*"daftar" + 0.078*"tidak" + 0.028*"data" + 0.023*"verifikasi" + 0.022*"email" + 0.022*"sesuai" + 0.019*"nomer" + 0.018*"akun" + 0.018*"pakai"
5	0.112*"tidak" + 0.066*"kode" + 0.066*"verifikasi" + 0.061*"otp" + 0.052*"nomor" + 0.051*"masuk" + 0.039*"sms" + 0.036*"kirin" + 0.035*"daftar" + 0.032*"pulsa"
6	0.065*"tidak" + 0.036*"verifikasi" + 0.031*"susah" + 0.030*"bikin" + 0.029*"wajah" + 0.028*"sulit" + 0.025*"nya" + 0.022*"banget" + 0.020*"gagal" + 0.020*"mudah"
7	0.087*"tidak" + 0.070*"login" + 0.053*"masuk" + 0.037*"daftar" + 0.034*"password" + 0.034*"salah" + 0.023*"nomor" + 0.022*"nik" + 0.021*"nya" + 0.020*"update"

Tabel 11 menyajikan 10 kata kunci utama yang muncul disertai dengan peluang kemunculannya pada tiap topik. Dari kata kunci utama ini dapat dilakukan interpretasi terkait judul bahasan tiap topik. Kata-kata dengan probabilitas yang tinggi menunjukkan fokus utama pada masing-masing topik, sehingga dapat memberikan informasi mengenai aspek yang dibahas. Hasil interpretasi tiap topiknya dapat dilihat pada Tabel 12.

Tabel 12. Hasil Interpretasi Topik

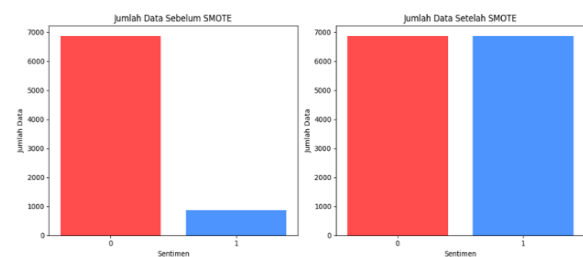
Topik	Nama Topik
0	Akses Membuka dan Menjalankan Aplikasi di Beberapa Perangkat
1	Jaringan dan Konektivitas Aplikasi
2	Fitur Administrasi (Pindah/Mengubah Fasilitas Kesehatan, Mengubah Data, dan Akses Kartu JKN)
3	Fitur Pendaftaran Berobat, Rujukan, Antrian, atau Pembayaran Online
4	Verifikasi, Daftar Akun, dan Pelayanan Administrasi Whatsapp
5	Kode Verifikasi OTP
6	Fitur Verifikasi Wajah
7	Masalah Login dan Kesalahan Data

4.5. Klasifikasi Sentimen

Klasifikasi sentimen dilakukan menggunakan algoritma *Random Forest*. Sebelum dilakukan pelatihan model, data terlebih dahulu diubah ke dalam bentuk representasi numerik menggunakan TF-IDF. Selanjutnya dilakukan pembagian data latih dan data uji, penanganan data tidak seimbang pada data latih, pelatihan model pada data latih, dan evaluasi model pada data uji.

Pada penelitian ini data dibagi menjadi 80% data latih dan 20% data uji. Karena data pada penelitian ini tidak seimbang untuk kelas negatif dan positifnya, maka dilakukan penanganan data tidak seimbang menggunakan SMOTE pada data latih. Perbandingan

data sebelum dan setelah dilakukan SMOTE dapat dilihat pada Gambar 4.



Gambar 4. Perbandingan Data Sebelum dan Setelah SMOTE

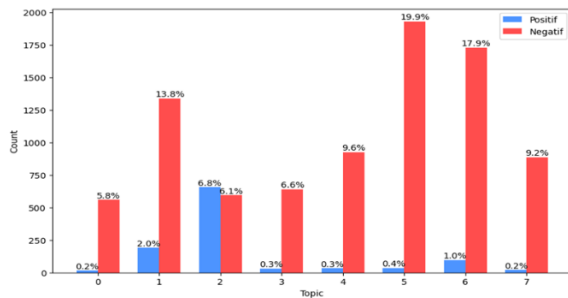
Berdasarkan Gambar 4 diketahui sebelum dilakukan SMOTE kelas negatif yang dilambangkan dengan 0 lebih banyak dibandingkan dengan kelas positif. Setelah dilakukan SMOTE, data dari kedua kelas menjadi seimbang. Kemudian, dilakukan evaluasi model pada data uji. Hasil evaluasinya dapat dilihat pada Tabel 13.

Tabel 13. Evaluasi Model

Kelas	Presisi	Recall	F1-Score	Akurasi
Positif	0.86	0.72	0.78	0.96
Negatif	0.97	0.99	0.98	

4.6. Interpretasi Sentimen dan Topik

Setelah topik-topik pada ulasan aplikasi Mobile JKN berhasil diidentifikasi dan klasifikasi sentimen telah dilakukan maka berikutnya akan dilakukan klasifikasi sentimen berdasarkan topik-topiknya. Hasil klasifikasi sentimen untuk tiap topiknya dapat dilihat pada Gambar 5.



Gambar 5. Distribusi Sentimen Berdasarkan Topik

Gambar 5 memperlihatkan sebaran sentimen di antara 8 topik yang teridentifikasi, yang masing-masing mewakili aspek-aspek spesifik dari pengalaman pengguna dengan aplikasi. Secara keseluruhan, ulasan dengan sentimen negatif mendominasi di sebagian besar topik, terutama pada topik 5 (kode verifikasi OTP) dan topik 6 (fitur verifikasi wajah) yaitu mencapai 19,9% dan 17,9% dari 9677 ulasan. Hal ini menunjukkan bahwa fitur kode verifikasi OTP dan fitur verifikasi wajah pada aplikasi Mobile JKN merupakan fitur yang paling banyak dikeluhkan pengguna. Topik 2 (fitur administrasi) merupakan satu-satunya topik dengan proporsi sentimen positif yang relatif tinggi, yaitu sebesar 6,8%, meskipun tetap disertai dengan sentimen negatif sebesar 6,1%. Ini mengindikasikan bahwa fitur administrasi yang tersedia dalam aplikasi Mobile JKN memiliki kesan beragam di mata pengguna, baik berupa kepuasan maupun kekecewaan. Sebaliknya, topik 0, topik 3, topik 4, dan topik 7 didominasi oleh sentimen negatif dengan kontribusi ulasan positif yang sangat kecil, yaitu 0,2 hingga 0,3%. Topik 1 juga didominasi oleh sentimen negatif, yaitu sebesar 13,8% dengan hanya 2% sentimen positif.

Secara keseluruhan, Gambar 5 mengindikasikan bahwa aspek verifikasi pengguna baik dalam bentuk kode OTP dan verifikasi wajah, login, serta konektivitas aplikasi menjadi perhatian utama pengguna dan menjadi sumber utama keluhan. Oleh karena itu, peningkatan dalam fitur OTP, verifikasi wajah, dan stabilitas aplikasi dapat menjadi fokus utama perbaikan untuk meningkatkan pengalaman pengguna Mobile JKN.

5. KESIMPULAN DAN SARAN

Metode *Latent Dirichlet Allocation* mengidentifikasi 8 topik utama pada 9.677 data ulasan aplikasi Mobile JKN dengan nilai *coherence score* sebesar 0,575. Topik kode verifikasi OTP memiliki proporsi ulasan negatif tertinggi yaitu 19,9%. Topik fitur administrasi menjadi topik dengan proporsi ulasan positif tertinggi yaitu sebesar 6,8%. Secara keseluruhan ulasan negatif masih mendominasi pada tiap topik sehingga diperlukan perbaikan fitur maupun layanan aplikasi Mobile JKN. Metode *Random Forest* yang digunakan dalam penelitian ini menunjukkan hasil yang sangat baik dalam mengklasifikasikan sentimen terhadap aplikasi Mobile JKN. Model ini memperoleh akurasi keseluruhan sebesar 96%.

Evaluasi untuk sentimen negatif, model menunjukkan presisi sebesar 97%, *recall* sebesar 99%, dan *F1-score* mencapai 98%. Untuk sentimen positif menunjukkan presisi sebesar 86% dan *recall* sebesar 72%, dengan *F1-score* mencapai 78%. Saran untuk penelitian berikutnya dapat mempertimbangan metode penanganan ketidakseimbangan kelas data yang lebih baik agar meningkatkan kualitas model pada kelas minoritas.

DAFTAR PUSTAKA

- [1] N. F. Alayda, C. M. Aulia, E. R. Ritonga, and S. H. Purba, "Literature Review: Analisis Dampak Kebijakan Jaminan Kesehatan Nasional (JKN) Terhadap Akses dan Kualitas Pelayanan Kesehatan," *Jurnal Kolaboratif Sains*, vol. 7, no. 7, pp. 2616–2626, Jul. 2024, doi: 10.56338/jks.v7i7.5573.
- [2] RSUD Ir. Soekarno Kabupaten Sukoharjo, "Duta Mobile JKN Sosialisasi Aplikasi Kepada Pengunjung," Duta Mobile JKN Sosialisasi Aplikasi Kepada Pengunjung.
- [3] Y. Kustiyahningsih and Y. Permana, "Penggunaan Latent Dirichlet Allocation (LDA) dan Support-Vector Machine (SVM) Untuk Menganalisis Sentimen Berdasarkan Aspek Dalam Ulasan Aplikasi EdLink," *Teknika*, vol. 13, no. 1, pp. 127–136, Mar. 2024, doi: 10.34148/teknika.v13i1.746.
- [4] D. A. Fitri and D. Damayanti, "KOMPARASI ALGORITMA RANDOM FOREST CLASSIFIER DAN SUPPORT VECTOR MACHINE UNTUK SENTIMEN MASYARAKAT TERHADAP PINJAMAN ONLINE DI MEDIA SOSIAL," *JIPi (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 9, no. 4, pp. 2018–2029, Nov. 2024, doi: 10.29100/jipi.v9i4.5608.
- [5] F. N. Hikmah, S. Basuki, and Y. Azhar, "Deteksi Topik Tentang Tokoh Publik Politik Menggunakan Latent Dirichlet Allocation," *REPOSITOR*, vol. 2, no. 4, pp. 415–426, Apr. 2020, [Online]. Available: <https://doi.org/10.22219/repositor.v2i4.30515>
- [6] E. Fitri, Y. Yuliani, S. Rosyida, and W. Gata, "Analisis Sentimen Terhadap Aplikasi Ruangguru Menggunakan Algoritma Naive Bayes, Random Forest Dan Support Vector Machine," *TRANSFORMATIKA*, vol. 18, no. 1, pp. 71–80, 2020, [Online]. Available: <https://doi.org/10.26623/transformatika.v18i1.2317>
- [7] E. Puspita, D. F. Shiddieq, and F. F. Roji, "Pemodelan Topik pada Media Berita Online Menggunakan Latent Dirichlet Allocation (Studi Kasus Merek Somethinc)," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 2, pp. 481–489, Apr. 2024, doi: 10.57152/malcom.v4i2.1204.

- [8] Mustafa, "Mobile JKN: Layanan dalam Satu Genggaman," Jaringan Inovasi Pelayanan Publik Nasional. Accessed: Feb. 10, 2025. [Online]. Available: <https://jippnas.menpan.go.id/inovasi/2649>
- [9] D. Alita and A. Rahman, "Pendeteksian Sarkasme pada Proses Analisis Sentimen Menggunakan Random Forest Classifier," 2020. [Online]. Available: <https://doi.org/10.23960/komputasi.v8i2.2615>
- [10] D. Nurwahidah, G. Dwilestari, N. D. Nuris, and R. Narasati, "ANALISIS SENTIMEN DATA ULASAN PENGGUNA APLIKASI GOOGLE KELAS PADA GOOGLE PLAY STORE MENGGUNAKAN ALGORITMA NAÏVE BAYES," 2023. [Online]. Available: <https://doi.org/10.36040/jati.v7i6.8245>
- [11] A. F. Anjani, D. Anggraeni, and I. M. Tirta, "Implementasi Random Forest Menggunakan SMOTE untuk Analisis Sentimen Ulasan Aplikasi Sister for Students UNEJ," *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 9, no. 2, pp. 163–172, Sep. 2023, doi: 10.25077/teknosi.v9i2.2023.163-172.
- [12] D. L. C. Pardede and M. A. I. Waskita, "ANALISIS PEMODELAN TOPIK UNTUK ULASAN TENTANG PEDULI LINDUNGI," *Jurnal Ilmiah Informatika Komputer*, vol. 28, no. 1, pp. 17–26, 2023, doi: 10.35760/ik.2023.v28i1.7925.
- [13] N. Royani, C. E. Widodo, and B. Warsito, "Topic Modelling Latent Dirichlet Allocation untuk Klasifikasi Komentar pada Layanan Streaming Platform," *JST (Jurnal Sains dan Teknologi)*, vol. 12, no. 3, Jan. 2024, doi: 10.23887/jstundiksha.v12i3.68492.