

## Analisis Sentimen dan Pemodelan Topik pada Ulasan Aplikasi AllStat BPS 2017-2025 dengan Pendekatan Latent Dirichlet Allocation

Gabriella Elisabeth Simanungkalit<sup>1</sup>, Hervira Nur Shafira<sup>2</sup>, Rani Nooraeni<sup>3</sup>

<sup>1,2</sup> Program Studi Statistika, Politeknik Statistika STIS, Indonesia

<sup>3</sup> Program Studi Komputasi Statistik, Politeknik Statistika STIS, Indonesia

<sup>1</sup>212212621@stis.ac.id, <sup>2</sup>212212642@stis.ac.id, <sup>3</sup>raninoor@stis.ac.id

### Info Artikel

#### Riwayat Artikel:

Received 2025-07-20

Revised 2026-02-03

Accepted 2026-02-10

**Abstract** – This research is motivated by the increasing need for statistical data that is easily accessible through mobile applications, one of which is the AllStat BPS application. The purpose of this study is to analyse sentiment and identify the main topics in user reviews of the AllStat BPS application in the period 2017–2025. This research contributes by providing a replicable, text-based analytical framework to support the evaluation and quality improvement of official statistical applications. The methods used include lexicon-based sentiment analysis with the InSet dictionary and classification using the Naive Bayes algorithm, Random Forest, and Support Vector Machine (SVM). Topic modelling was carried out using the Latent Dirichlet Allocation (LDA) approach. The results showed that the Random Forest model provided the best classification performance with an accuracy on training data of 88.16% and a kappa value of 0.8046. In addition, LDA successfully identified eight main topics from user reviews, with Topic 1 having the highest coherence value (0.1784) indicating the semantic strength between words in the topic. These topics were then discussed into a software quality framework based on the ISO/IEC 25010 standard, with the aspects of Functional Suitability and Performance Efficiency as the dominant topics. The conclusion of this study is that the combination of Random Forest and LDA methods is effective in classifying sentiments and describing the focus of issues in user reviews of the BPS AllStat application.

**Keywords:** AllStat BPS, LDA, Naïve Bayes, Random Forest, SVM

#### Corresponding Author:

Rani Nooraeni

Email: raninoor@stis.ac.id



This is an open access article under the [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/) license.

**Abstrak** – Penelitian ini dilatarbelakangi oleh meningkatnya kebutuhan akan data statistik yang mudah diakses melalui aplikasi mobile, salah satunya adalah aplikasi AllStat BPS. Tujuan dari penelitian ini adalah untuk menganalisis sentimen dan mengidentifikasi topik utama dalam ulasan pengguna aplikasi AllStat BPS pada periode 2017–2025. Penelitian ini berkontribusi dengan menyediakan kerangka analitis berbasis data teks yang dapat direplikasi untuk mendukung evaluasi dan peningkatan kualitas aplikasi statistik resmi. Metode yang digunakan mencakup analisis sentimen berbasis lexicon dengan kamus InSet dan klasifikasi menggunakan algoritma Naive Bayes, Random Forest, dan Support Vector Machine (SVM). Pemodelan topik dilakukan dengan pendekatan Latent Dirichlet Allocation (LDA). Hasil penelitian menunjukkan bahwa model Random Forest memberikan performa klasifikasi terbaik dengan akurasi pada data latih sebesar 88,16% dan nilai kappa 0,8046. Selain itu, LDA berhasil mengidentifikasi delapan topik utama dari ulasan pengguna, dengan Topik 1 memiliki nilai koherensi tertinggi (0,1784) yang mengindikasikan kekuatan semantik antar kata dalam topik tersebut. Topik-topik ini kemudian dipetakan ke dalam kerangka kualitas perangkat lunak berdasarkan standar ISO/IEC 25010, dengan aspek Functional Suitability dan Performance Efficiency sebagai topik dominan. Kesimpulan dari penelitian ini adalah bahwa kombinasi metode Random Forest dan LDA efektif dalam mengklasifikasikan sentimen serta menggambarkan fokus isu dalam ulasan pengguna aplikasi AllStat BPS.

**Kata Kunci:** AllStat BPS, LDA, Naïve Bayes, Random Forest, SVM

## I. PENDAHULUAN

Di era transformasi digital global, berbagai organisasi internasional telah meluncurkan aplikasi berbasis mobile untuk menyediakan data statistik secara terbuka dan cepat. United Nations Statistics Division (UNSD) adalah lembaga Perserikatan Bangsa-Bangsa (PBB) yang mengatur dan memberi standar internasional dalam pengumpulan dan penyajian data statistik global. UNSD meluncurkan aplikasi UNdata sebagai bentuk komitmen mereka dalam menyediakan akses terbuka terhadap data global. Aplikasi ini menyajikan indikator kunci sosial, ekonomi, dan lingkungan dari lebih dari 200 negara secara terstruktur dan dapat diakses secara portabel [1].

Kebutuhan akan data yang cepat, akurat, dan kredibel semakin meningkat seiring dengan pentingnya pengambilan keputusan berbasis data dalam berbagai sektor. Negara-negara dengan sistem data yang kuat umumnya lebih siap menghadapi tantangan seperti krisis ekonomi, masalah kesehatan, dan persoalan sosial lainnya [2]. Organisasi PBB melalui program Global SDG Indicators Database juga menyatakan bahwa data terbuka berperan penting dalam memantau pencapaian tujuan pembangunan berkelanjutan (SDGs) [3].

Di Indonesia, transformasi digital juga menjadi pendorong utama meningkatnya kebutuhan masyarakat akan data statistik yang instan dan praktis. Badan Pusat Statistik (BPS) merespons hal ini dengan mengembangkan aplikasi Allstats BPS, yang memungkinkan masyarakat mengakses berbagai produk BPS seperti publikasi, indikator strategis, dan infografis secara langsung melalui perangkat seluler [4]. Berdasarkan Survei Kebutuhan Data BPS tahun 2022, sekitar 45,49% pengguna mengandalkan media digital, termasuk situs web dan aplikasi Allstats, sebagai sumber utama informasi statistik [5]. Sejak peluncurannya pada 2017, aplikasi ini telah diunduh lebih dari 100.000 kali, dan mendapatkan rating 4,3 hingga 4,5 di Google Play Store dan App Store [6].

Meskipun demikian, ulasan pengguna menunjukkan berbagai persepsi. Sebagian besar memberikan penilaian positif atas kemudahan akses terhadap data statistik dan kelengkapan informasi, namun sejumlah pengguna menyampaikan keluhan seperti keterbatasan fitur pencarian dan kesulitan mengunduh publikasi [7]. Ulasan-ulasan tersebut mencerminkan pengalaman pengguna secara langsung dan berpotensi mengandung informasi penting yang belum sepenuhnya dimanfaatkan sebagai bahan evaluasi pengembangan aplikasi. Oleh karena itu, diperlukan pendekatan analitis yang sistematis untuk menggali pola persepsi pengguna secara lebih mendalam.

Untuk mengidentifikasi persepsi tersebut, peneliti menggunakan beberapa metode klasifikasi seperti Naive Bayes, Random Forest, dan Support Vector Machine (SVM). Dalam konteks aplikasi Allstats BPS, penelitian sebelumnya berhasil mengelompokkan komentar menjadi positif dan negatif dengan tingkat akurasi klasifikasi menggunakan Naive Bayes sebesar 86,83% [7]. Metode Naive Bayes Classifier digunakan karena keefektifannya dalam klasifikasi teks dan kemampuannya dalam menangani dataset besar dengan akurasi yang tinggi [8].

Salah satu penelitian yang membandingkan ketiga metode klasifikasi menghasilkan akurasi metode Naive Bayes, Random Forest, dan SVM secara berturut-turut sebesar 94.16%, 97.16%, dan 96.01% [9]. Penelitian lain menggunakan metode Naive Bayes dan SVM untuk menganalisis sentimen aplikasi Peluang menghasilkan nilai akurasi SVM lebih baik dibanding Naive Bayes sebesar 99.5% [10]. Model Random Forest menghasilkan nilai akurasi yang lebih baik dibandingkan metode Naive Bayes dalam menganalisis sentimen positif pada aplikasi getcontact sebesar 80%, namun pada sentimen negatif nilai akurasi kedua metode memberikan hasil serupa [11]. Dikarenakan terbatasnya model yang sudah diteliti untuk analisis sentimen pada aplikasi Allstat BPS, peneliti mengambil beberapa model terbaik yang sudah banyak diteliti pada analisis sentimen aplikasi lainnya. Hal ini bertujuan untuk menemukan model analisis sentimen terbaik untuk penelitian ini.

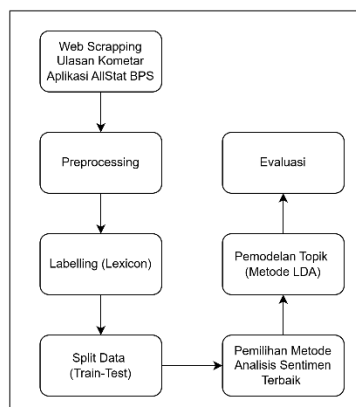
Selain analisis sentimen, peneliti juga ingin melakukan klasifikasi topik pada hasil sentimen yang diperoleh untuk menemukan topik-topik yang menjadi fokus pada ulasan aplikasi Allstat BPS. Hingga saat ini, penelitian terkait aplikasi Allstats BPS masih berfokus pada klasifikasi sentimen dan belum mengkaji struktur topik yang muncul dalam ulasan pengguna. Salah satu pendekatan terbaik dan banyak digunakan untuk pemodelan topik adalah Latent Dirichlet Allocation (LDA). Suatu studi menunjukkan bahwa metode LDA mampu mengidentifikasi topik secara efektif pada data teks ulasan [12]. Penelitian sebelumnya menerapkan pemodelan topik pada komentar YouTube menggunakan metode LDA dan K-Means dengan representasi teks TF-IDF dan GloVe untuk mengidentifikasi topik-topik dominan dalam diskusi publik. Hasil penelitian menunjukkan bahwa LDA dengan TF-IDF menghasilkan nilai koherensi tertinggi, sementara K-Means dengan GloVe 100D memberikan interpretabilitas topik yang lebih baik, sehingga menunjukkan adanya trade-off antara koherensi dan keterbacaan topik [13].

Berdasarkan uraian tersebut, penelitian ini memiliki kebaruan dalam menggabungkan analisis sentimen dan pemodelan topik menggunakan pendekatan LDA pada ulasan aplikasi layanan publik, khususnya aplikasi AllStat BPS. Pendekatan ini memperkaya metode analisis teks dalam konteks aplikasi pemerintah yang masih relatif terbatas dibandingkan aplikasi komersial. Kontribusi penelitian ini tidak hanya terletak pada penentuan metode klasifikasi sentimen yang paling optimal, tetapi juga pada pemetaan topik utama ulasan pengguna sebagai bahan evaluasi kualitas aplikasi. Dengan demikian, tujuan penelitian ini adalah untuk membandingkan beberapa metode klasifikasi dalam menganalisis sentimen serta membentuk pemodelan topik untuk mengidentifikasi klasifikasi topik yang muncul pada ulasan pengguna terhadap aplikasi Allstats BPS di Google Play Store selama periode 2017–2025. Hasil penelitian ini diharapkan dapat memberikan kontribusi empiris bagi pengembangan metode analisis sentimen dalam aplikasi statistik, sekaligus menjadi dasar rekomendasi bagi BPS dalam upaya meningkatkan kualitas layanan aplikasi Allstats BPS di masa mendatang.

## II. METODE

Data yang digunakan dalam penelitian ini berupa ulasan pengguna aplikasi Allstats BPS yang diperoleh melalui teknik *web scraping* dari Google Play Store menggunakan bahasa pemrograman Python, dengan rentang waktu 2017–2025. Dengan menggunakan data ulasan pengguna dalam rentang waktu yang panjang (2017–2025), penelitian ini memberikan kontribusi berupa analisis dinamika sentimen dan perubahan topik dari waktu ke waktu, sehingga mampu menggambarkan evolusi persepsi pengguna terhadap kualitas layanan digital BPS secara berkelanjutan. Setelah proses pengumpulan data, diperoleh 669 ulasan yang digunakan sebagai objek penelitian yang kemudian dikategorikan ke dalam tiga jenis sentimen, yaitu positif, netral, dan negatif. Pelabelan sentimen tersebut dilakukan dengan menggunakan keyword kamus lexicon pada software python.

Tahapan penelitian ini dirancang secara sistematis untuk memastikan transparansi dan reproduktibilitas hasil, mulai dari pengumpulan data hingga evaluasi model topik. Alur lengkap prosedur penelitian disajikan secara visual dalam Gambar 1.



Gambar 1. Tahap Penelitian

Tahap awal dalam pengolahan data adalah preprocessing, yang meliputi pembersihan data dari karakter tidak relevan, proses tokenization, pelabelan sentimen, serta pembagian data menjadi data latih dan data uji [14]. Tahapan preprocessing dilakukan untuk membersihkan serta mempersiapkan data ulasan pengguna aplikasi AllStat BPS agar layak digunakan dalam analisis sentimen dan pemodelan topik. Proses ini mencakup serangkaian langkah terstruktur, termasuk penghapusan data duplikat dan baris yang mengandung nilai kosong pada kolom ulasan, guna menjamin kualitas data yang dianalisis.

Selanjutnya, dilakukan proses pembersihan teks, yang mencakup konversi seluruh karakter ke huruf kecil (*lowercasing*), penghapusan angka, tanda baca, karakter non-alfabet, serta spasi berlebih. Setelah itu, teks diubah menjadi token atau potongan kata menggunakan metode tokenisasi. Proses ini diikuti dengan penghapusan stopword, yaitu kata-kata umum dalam Bahasa Indonesia yang tidak memiliki kontribusi bermakna terhadap analisis, seperti "yang", "dan", "di", dan sebagainya. Daftar stopwords yang digunakan merupakan kombinasi dari pustaka NLTK dan kamus Bahasa Indonesia dari Sastrawi.

Langkah selanjutnya adalah melakukan proses stemming dengan menggunakan stemmer Bahasa Indonesia dari pustaka Sastrawi. Tujuan dari proses ini adalah untuk mengembalikan setiap kata ke bentuk dasarnya, seperti kata "meningkatkan" yang akan diubah menjadi "tingkat". Setelah melalui tahap tokenisasi, stopwords removal, dan stemming, hasilnya kemudian direkonstruksi menjadi kalimat yang telah dibersihkan (*cleaned review*).

Setelah data melalui tahap *preprocessing*, dilakukan transformasi untuk mengubah data teks ke dalam bentuk numerik menggunakan teknik *feature extraction*, sehingga dapat diproses oleh algoritma klasifikasi [15]. Pelabelan kelas sentimen pada ulasan pengguna aplikasi AllStat BPS dilakukan dengan menggunakan pendekatan berbasis leksikon, yaitu dengan memanfaatkan daftar kata berbahasa Indonesia yang telah diklasifikasikan ke dalam kategori positif dan negatif. Dalam penelitian ini, digunakan InSet (Indonesia Sentiment Lexicon), yaitu kamus sentimen yang dikembangkan oleh Fajri Koto dan tim, yang secara khusus disusun untuk mendukung analisis sentimen dalam konteks bahasa Indonesia. InSet secara keseluruhan memuat 3.609 entri kata, yang terbagi dalam tiga kategori utama: positif, negatif, dan netral. Dalam pendekatan ini, digunakan kamus yang berisi daftar kata positif dan negatif, sementara kata-kata yang tidak termasuk dalam kedua kategori tersebut diklasifikasikan sebagai netral, berdasarkan skor polaritas (*polarity score*) yang diperoleh. [16].

Kata-kata dari kedua kategori tersebut dipisahkan ke dalam dua file eksternal dan digunakan untuk proses pelabelan otomatis terhadap data ulasan. Setiap ulasan dianalisis berdasarkan jumlah kata yang cocok dengan daftar positif maupun negatif tersebut. Skor polaritas kemudian diberikan dengan aturan: ulasan diberi skor +1 apabila mengandung lebih banyak kata positif, -1 apabila mengandung lebih banyak kata negatif, dan 0 jika tidak ditemukan kata dari kedua kategori tersebut. Untuk keperluan klasifikasi sentimen, dataset yang telah berlabel dibagi (*data splitting*) dengan proporsi 80% sebagai data latih (*training data*) dan 20% sebagai data uji (*testing data*) menggunakan teknik *stratified sampling*. Ekstraksi fitur dilakukan menggunakan metode TF-IDF (*Term Frequency-Inverse Document Frequency*).

Penelitian ini memiliki dua tujuan utama, yaitu melakukan analisis sentimen terhadap ulasan pengguna aplikasi AllStat BPS serta mengidentifikasi topik-topik utama yang muncul melalui pendekatan pemodelan topik. Kontribusi ini memberikan pemahaman yang lebih mendalam mengenai faktor-faktor utama yang memicu sentimen positif maupun negatif dalam ulasan pengguna aplikasi statistik pemerintah. Penelitian terdahulu yang melakukan analisis sentimen terhadap aplikasi AllStat BPS menggunakan model Naive Bayes. Karena terbatasnya rujukan model terbaik

untuk analisis sentimen aplikasi AllSat BPS, peneliti melakukan komparasi model. Dalam penelitian ini, digunakan tiga jenis metode klasifikasi yaitu Naive Bayes, Random Forest, dan Support Vector Machine (SVM) yang masing-masing memiliki karakteristik berbeda dalam mengolah dan mengklasifikasikan data teks.

Model klasifikasi dibangun menggunakan Naive Bayes Classifier, yang bekerja berdasarkan prinsip probabilitas untuk menentukan kategori sentimen dari tiap komentar. Proses klasifikasi dilakukan dalam dua tahap, yaitu training untuk membangun model berdasarkan data berlabel, dan testing untuk menguji akurasi prediksi terhadap data uji [17]. Metode Random Forest Classifier merupakan teknik klasifikasi yang terdiri dari sejumlah pohon keputusan, di mana hasil akhir ditentukan melalui proses voting dari masing-masing pohon. Setiap pohon dilatih menggunakan data latih dan kumpulan fitur acak yang saling independen dan berbeda satu sama lain [18]. Metode SVM digunakan dalam penelitian ini karena kemampuannya dalam memisahkan kelas sentimen secara optimal pada data teks berdimensi tinggi. Metode ini bekerja dengan cara menentukan hyperplane optimal yang mampu memaksimalkan pemisahan antar kelas yang berbeda [19]. Nantinya, untuk mengukur kinerja ketiga model, penelitian ini menggunakan matrik evaluasi berupa akurasi, presisi, recall, dan F1-score, yang secara keseluruhan memberikan gambaran objektif tentang efektivitas model dalam mengelompokkan sentimen [20].

Tujuan kedua dari penelitian ini adalah merancang pemodelan topik yang dapat dijadikan dasar dalam mengevaluasi aplikasi Allstat. Dalam proses analisis topik, digunakan pendekatan Latent Dirichlet Allocation (LDA) sebagai metode utama penelitian ini. Metode LDA digunakan dalam penelitian ini untuk mengelompokkan kata-kata dalam ulasan pengguna ke dalam sejumlah topik berdasarkan pola distribusi probabilitas kemunculan kata. Latent Dirichlet Allocation (LDA) merupakan model probabilistik dalam topic modelling yang digunakan pada data teks untuk mengekstraksi informasi berupa topik-topik laten. Model LDA mampu melakukan reduksi dimensi representasi teks ke dalam ruang semantik serta memodelkan dokumen berdasarkan probabilitas kemunculan kosa kata, sehingga dapat mengurangi kompleksitas dan permasalahan data hingga batas tertentu [21]. Penelitian ini mengadopsi tahapan kerja metode LDA sebagaimana dijelaskan oleh Merawati et al. (2021), yang mencakup pembuatan kamus dan korpus dari data sentimen, penentuan kata-kata untuk topik tertentu berdasarkan distribusi Dirichlet, perhitungan probabilitas kata pada tiap topik, dan pengulangan proses tersebut hingga seluruh kata dalam korpus dianalisis [12].

### III. HASIL DAN PEMBAHASAN

Dari proses pengumpulan data menggunakan teknik *web scraping* pada tanggal 12 Juni 2025 total ulasan yang berhasil diperoleh adalah 669 data ulasan. Persebaran jumlah berdasarkan tahun ditunjukkan pada Gambar 2.



Gambar 2. Jumlah Ulasan Aplikasi AllStat BPS per Tahun

Dari Gambar 2 dapat diketahui bahwa dari tahun 2017 hingga ulasan terakhir pada 3 Juni 2025, jumlah ulasan tertinggi terjadi pada tahun 2019 kemudian menurun pada tahun 2020 hingga 2023. Pola ini mengindikasikan adanya perubahan tingkat partisipasi pengguna dalam memberikan umpan balik terhadap aplikasi seiring waktu.

Tahap selanjutnya adalah seleksi atribut dan pra-pemrosesan teks. Proses pra-pemrosesan meliputi *case folding*, *cleaning*, *normalization*, *filtering*, dan *stemming*. Tahapan ini bertujuan untuk mengurangi *noise* dan meningkatkan kualitas representasi teks sebelum dilakukan analisis lanjutan. Dari total 669 ulasan awal, sebanyak 430 ulasan dinyatakan valid dan layak digunakan dalam proses analisis setelah pra-pemrosesan. Tabel 1 berikut merupakan contoh data sebelum dan sesudah dilakukan preprocessing.

TABEL 1  
PERBANDINGAN DATA SEBELUM DAN SESUDAH TAHAPAN PREPROCESSING

Sebelum Preprocessing	Setelah Preprocessing
Aplikasi ini memudahkan pengguna mengakses informasi yg di cari sesuai kebutuhan yang ingin di dapat dari berbagai aspek dan bidang secara berimbang, serta bisa di jadikan barometer atau referensi dalam memperoleh informasi data yang akurat dan ter update.. ini merupakan wujud kolaborasi yg baik antar lini dan pemangku kebijakan untuk mewujudkan kepentingan bersama tentang potensi serta solusi masalah dlm berbagai bidang tentang keterbukaan informasi publik. semoga BPS semakin maju, kreatif Go BPS	['aplikasi', 'mudah', 'guna', 'akses', 'informasi', 'cari', 'sesuai', 'butuh', 'aspek', 'bidang', 'imbang', 'jadi', 'barometer', 'referensi', 'oleh', 'informasi', 'data', 'akurat', 'ter', 'baru', 'wujud', 'kolaborasi', 'lini', 'mang', 'bijak', 'wujud', 'penting', 'potensi', 'solusi', 'bidang', 'buka', 'informasi', 'publik', 'moga', 'bps', 'maju', 'kreatif', 'go', 'bps']

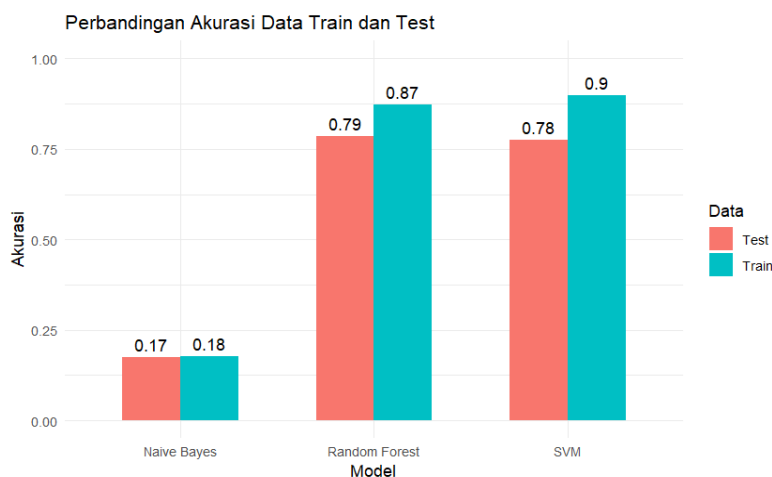
Pelabelan sentimen pada penelitian ini dilakukan menggunakan pendekatan *lexicon-based* dengan memanfaatkan kamus sentimen bahasa Indonesia yang dikembangkan oleh Fajri Koto et al. Pendekatan ini mengklasifikasikan sentimen ulasan ke dalam tiga kategori, yaitu positif, negatif, dan netral, berdasarkan akumulasi skor polaritas kata [16]. Pelabelan dilakukan melalui pemrograman Python dengan menghasilkan klasifikasi sentimen terhadap data yang dapat dilihat pada Tabel 2.

TABEL 2  
JUMLAH DATA HASIL PELABELAN LEXICON-BASED

Sentimen	Jumlah	Persentase (%)
Negatif	134	31.16
Neutral	47	10.93
Positif	249	57.91

Hasil pelabelan menunjukkan bahwa sentimen positif mendominasi dengan persentase sebesar 57,91%, diikuti oleh sentimen negatif sebesar 31,16%, dan sentimen netral sebesar 10,93%. Dominasi sentimen positif mengindikasikan bahwa secara umum pengguna menilai aplikasi AllStat BPS memberikan manfaat dalam penyediaan informasi statistik. Namun, proporsi sentimen negatif yang relatif besar juga menunjukkan adanya permasalahan yang dirasakan pengguna, terutama terkait performa dan keandalan sistem.

Penelitian ini menguji tiga algoritma klasifikasi sentimen, yaitu Naive Bayes, Random Forest, dan Support Vector Machine (SVM). Pada Tabel 3, tertera perbandingan untuk data train dan test untuk ketiga model, baik akurasi, kappa, maupun balanced accuracy-nya.



Gambar 3. Grafik Perbandingan Akurasi Metode Klasifikasi

TABEL 3  
PERBANDINGAN TIGA ALGORITMA METODE KLASIFIKASI SENTIMEN

Model	Dataset	Akurasi (%)	Kappa	Balanced Accuracy
Naive Bayes	Train	17.63	0.0055	0.5041
Naive Bayes	Test	17.35	0.0108	0.5082
Random Forest	Train	88.16	0.8046	0.9107
Random Forest	Test	79.59	0.6691	0.8793
SVM	Train	89.92	0.8306	0.9160
SVM	Test	77.55	0.6236	0.8186

Hasil pengujian menunjukkan bahwa algoritma Naive Bayes menghasilkan akurasi yang rendah, yakni sekitar 17%, dengan nilai kappa yang mendekati nol. Hal ini mengindikasikan bahwa algoritma tersebut kurang andal dalam melakukan klasifikasi sentimen. Sebaliknya, algoritma Random Forest dan SVM menunjukkan kinerja yang jauh lebih baik, dengan akurasi melebihi 77% pada data uji. Di antara ketiga model tersebut, Random Forest dipilih sebagai model terbaik karena tidak hanya memiliki akurasi dan balanced accuracy tertinggi, tetapi juga menunjukkan performa yang konsisten antara data latih dan data uji. Konsistensi ini menandakan bahwa model Random Forest lebih stabil dan tidak mengalami overfitting. Hasil ini sejalan dengan penelitian pada data ulasan film yang menunjukkan bahwa metode Random Forest menghasilkan akurasi tertinggi sebesar 88.8% dibandingkan Naive Bayes dan SVM dalam klasifikasi sentimen [22].

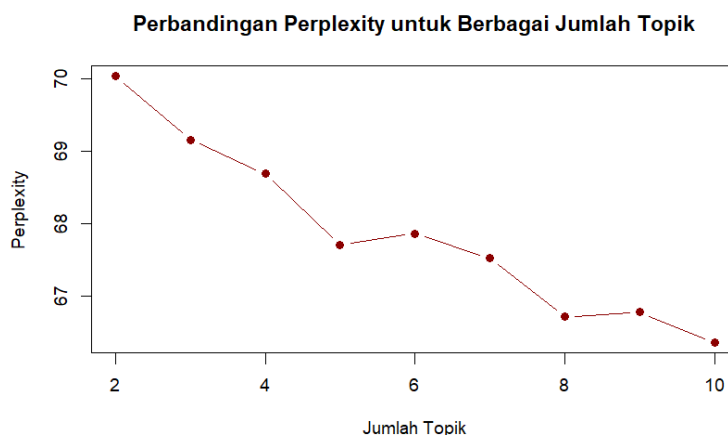
Random Forest dipilih sebagai model terbaik karena menunjukkan keseimbangan antara akurasi, *balanced accuracy*, dan stabilitas performa antara data latih dan data uji. Nilai Cohen's Kappa yang tinggi mencerminkan tingkat kesepakatan yang kuat antara hasil prediksi model dan label aktual. Temuan ini sejalan dengan berbagai penelitian sebelumnya yang menunjukkan bahwa algoritma Random Forest memiliki kinerja unggul dalam menangani data teks berdimensi tinggi serta distribusi kelas yang tidak seimbang.

TABEL 4  
CONFUSION MATRIX METODE RANDOM FOREST

Prediction	Reference		
	Negatif	Netral	Positif
Negatif	100	11	8
Netral	3	57	19
Positif	6	0	193

Berdasarkan hasil evaluasi menggunakan *confusion matrix*, model Random Forest menunjukkan kinerja klasifikasi yang baik pada data uji. Dari total 397 ulasan, model mampu mengklasifikasikan secara benar sebanyak 350 ulasan, yang terdiri dari 100 ulasan negatif, 57 ulasan netral, dan 193 ulasan positif. Hasil ini menghasilkan tingkat akurasi keseluruhan sebesar 88,16%, yang menunjukkan bahwa mayoritas sentimen pengguna berhasil diprediksi dengan tepat oleh model.

Penentuan jumlah topik optimal dalam pemodelan topik menggunakan pendekatan LDA dilakukan dengan mempertimbangkan nilai perplexity. Strategi pemilihan jumlah topik berdasarkan nilai perplexity ini sejalan dengan praktik umum pada penelitian pemodelan topik menggunakan LDA, di mana nilai perplexity digunakan sebagai indikator model fit untuk menentukan jumlah topik optimal [23]. Untuk itu, dilakukan perbandingan nilai perplexity pada berbagai jumlah topik ( $k$ ), mulai dari 2 hingga 10. Hasil visualisasi menunjukkan bahwa nilai perplexity cenderung menurun seiring bertambahnya jumlah topik, yang menandakan adanya peningkatan kualitas model.

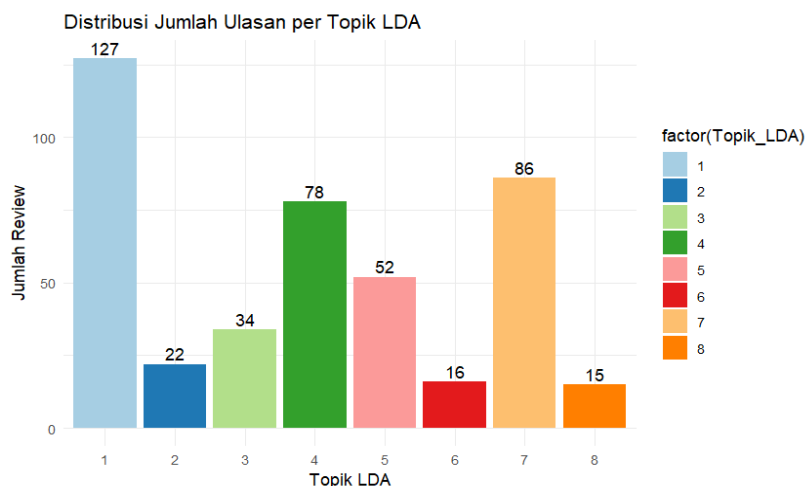


Gambar 4. Perbandingan Perplexity untuk Berbagai Jumlah Topik

Terlihat pada Gambar 4, penurunan perplexity yang paling signifikan terjadi hingga jumlah topik mencapai  $k = 8$ , sementara setelahnya penurunan nilai perplexity menjadi lebih landai. Berdasarkan pola tersebut, dipilih jumlah topik sebanyak delapan ( $k = 8$ ) sebagai nilai optimal, karena dinilai memberikan keseimbangan yang baik antara kompleksitas model dan kemudahan interpretasi hasil.

Setelah diperoleh jumlah topik optimal sebanyak  $k = 8$  berdasarkan evaluasi nilai perplexity, analisis topik dilanjutkan dengan menerapkan pemodelan Latent Dirichlet Allocation (LDA) pada data ulasan pengguna aplikasi AllStat BPS. Algoritma LDA terbukti mampu mengidentifikasi topik-topik utama dari ulasan pengguna aplikasi

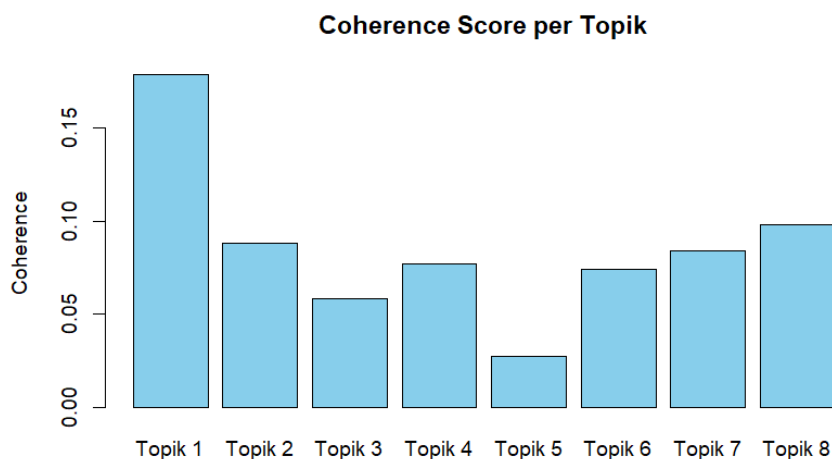
digital secara efektif, menghasilkan nilai koherensi yang tinggi serta topik yang mudah diinterpretasikan, sehingga memberikan gambaran yang jelas mengenai persepsi pengguna terhadap aplikasi[24].



Gambar 5. Distribusi Jumlah Ulasan per Topik LDA

Hasil distribusi topik dari model LDA pada Gambar 5 menunjukkan bahwa topik ke-1 merupakan topik yang paling dominan, dengan jumlah ulasan sebanyak 127, diikuti oleh topik ke-7 dengan 86 ulasan, dan topik ke-4 dengan 78 ulasan. Sementara itu, topik ke-6 dan ke-8 merupakan topik dengan jumlah ulasan paling sedikit, masing-masing hanya 16 dan 15 ulasan. Distribusi ini mengindikasikan bahwa ada beberapa tema utama yang menjadi perhatian utama pengguna, sementara topik lainnya hanya dibahas oleh sebagian kecil pengguna. Temuan ini dapat menjadi dasar bagi pengembang untuk menelusuri lebih lanjut isi dari masing-masing topik guna memahami fokus perhatian, keluhan, maupun kepuasan pengguna terhadap fitur atau performa aplikasi AllStat BPS.

Nilai koherensi digunakan untuk menilai seberapa kuat hubungan semantik antar kata dalam suatu topik, sehingga menjadi indikator penting dalam mengevaluasi kualitas hasil pemodelan LDA.



Gambar 6. Grafik Nilai Koheren per Topik

TABEL 5  
 RINCIAN NILAI KOHEREN PER TOPIK

Topik	Topik 1	Topik 2	Topik 3	Topik 4	Topik 5	Topik 6	Topik 7	Topik 8
Coherence Score	0.1784	0.0881	0.0586	0.0771	0.0277	0.0739	0.0843	0.0978

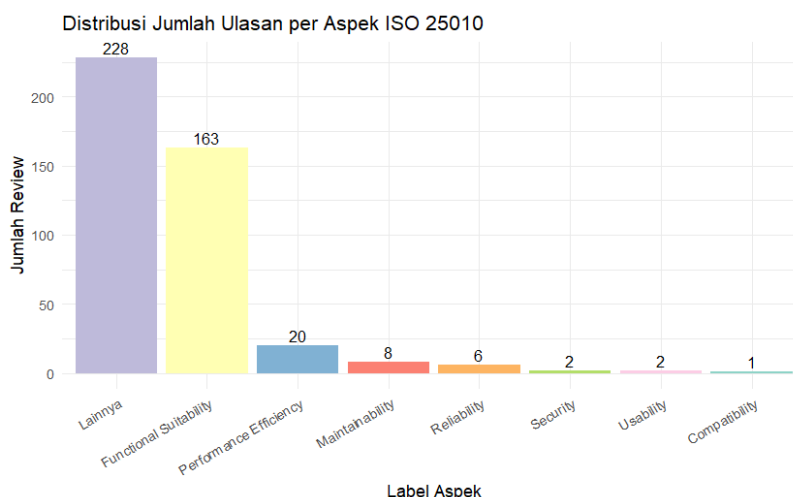
Berdasarkan hasil analisis pada Gambar 6 dan Tabel 5, Topik 1 menunjukkan nilai koherensi tertinggi sebesar 0.1784, yang menandakan bahwa kata-kata dalam topik tersebut memiliki konsistensi makna yang paling tinggi dibandingkan dengan topik lainnya. Sebaliknya, Topik 5 memperoleh nilai koherensi terendah, yaitu 0.0277, yang menunjukkan lemahnya keterkaitan semantik di dalamnya, sehingga topik ini dinilai kurang representatif. Nilai

koherensi pada topik-topik lainnya berkisar antara 0.0586 hingga 0.0978, dengan sebagian besar berada di bawah angka 0.1. Informasi tersebut, sebagaimana disajikan dalam Gambar 6 dan Tabel 5, menjadi dasar untuk mengevaluasi kualitas masing-masing topik serta mempertimbangkan kemungkinan perbaikan model, baik melalui penyesuaian jumlah topik maupun parameter lainnya.

TABEL 6  
 PELABELAN TOPIK DENGAN STANDAR ISO UNTUK REKAYASA PERANGKAT LUNAK

Aspek ISO 25010	Keyword Utama (ID & EN)
Functional Suitability	fitur, fungsi, data, informasi, statistik, tidak lengkap, tidak akurat, akurat, lengkap, feature, function, complete, correct, accurate
Performance Efficiency	lemot, lambat, cepat, berat, ringan, crash, loading, delay, respon, ngelag, slow, fast, freeze, lag
Compatibility	kompatibel, tidak cocok, tidak bisa dibuka, versi android, device, OS, compatible, android version, support
Usability	mudah digunakan, user friendly, tampilan, UI, navigasi, desain, antarmuka, susah digunakan, interface, design, easy to use, layout
Reliability	error, gagal, bug, force close, sering berhenti, tidak stabil, hang, glitch, unstable, crashed
Security	aman, keamanan, privasi, data bocor, tidak aman, akses ilegal, secure, privacy, hacked
Maintainability	update, pembaruan, perbaikan, pemeliharaan, bug fix, maintenance, fix, patch
Portability	multiplatform, bisa di ios, bisa di semua hp, di semua os, kompatibel semua perangkat, cross-platform, multi-device

Tabel 6 menyajikan pelabelan topik berdasarkan standar ISO/IEC 25010, yang umum digunakan dalam mengevaluasi kualitas perangkat lunak. Delapan aspek utama yang digunakan mencakup: Functional Suitability (fitur, fungsi, data, informasi), Performance Efficiency (lemot, cepat, loading, respon), Compatibility (versi perangkat dan sistem operasi), Usability (kemudahan penggunaan dan antarmuka), Reliability (stabilitas sistem seperti error dan bug), Security (privasi dan akses ilegal), Maintainability (pembaruan dan perbaikan), serta Portability (dukungan lintas perangkat dan platform). Kerangka ini memberikan dasar sistematis dalam mengelompokkan topik ulasan pengguna ke dalam dimensi kualitas perangkat lunak. Penggunaan standar ISO/IEC 25010 dalam evaluasi kualitas aplikasi juga telah diterapkan dalam berbagai penelitian terdahulu, seperti pada studi yang mengevaluasi kualitas aplikasi Menu Harianku menggunakan pendekatan ISO/IEC 25010:2011 [25]. Dengan demikian, tabel ini memberikan acuan sistematis dalam mengelompokkan kata-kata sentimen atau topik ulasan pengguna ke dalam aspek kualitas perangkat lunak sesuai standar ISO.



Gambar 7. Distribusi Jumlah Ulasan per Topik (Standar ISO 25010)

Distribusi jumlah ulasan berdasarkan aspek ISO 25010 dalam Gambar 7 menunjukkan bahwa sebagian besar ulasan pengguna jatuh ke dalam kategori "Lainnya" dengan total 228 ulasan, yang kemungkinan mencakup konten yang tidak secara eksplisit sesuai dengan delapan aspek standar ISO. Aspek Functional Suitability menjadi kategori ISO yang paling dominan, dengan 163 ulasan, diikuti oleh Performance Efficiency sebanyak 20 ulasan. Sementara itu, aspek lain seperti Maintainability (8 ulasan), Reliability (6 ulasan), Security dan Usability (masing-masing 2 ulasan), serta Compatibility (1 ulasan) memiliki distribusi yang jauh lebih rendah. Temuan ini menunjukkan bahwa



Dibandingkan dengan penelitian terdahulu, hasil penelitian ini konsisten dalam menunjukkan keunggulan Random Forest untuk klasifikasi sentimen berbasis teks. Penggunaan LDA dengan evaluasi perplexity juga sejalan dengan praktik umum dalam literatur text mining, meskipun nilai koherensi yang relatif rendah pada beberapa topik menegaskan keterbatasan metode ini pada data ulasan pendek.

Keberhasilan penelitian ini didukung oleh integrasi beberapa metode analisis dan penggunaan standar ISO/IEC 25010 sebagai kerangka evaluasi kualitas perangkat lunak. Namun demikian, penelitian ini memiliki keterbatasan, antara lain jumlah data ulasan yang terbatas, penggunaan pelabelan sentimen berbasis leksikon yang berpotensi menimbulkan bias, serta subjektivitas dalam pemetaan topik ke aspek ISO/IEC 25010. Melalui pengaitan hasil pemodelan topik LDA dengan polaritas sentimen, penelitian ini mengidentifikasi topik-topik kunci yang secara signifikan berkontribusi terhadap sentimen positif, negative, dan netral pengguna. Temuan ini memberikan kontribusi konseptual dalam memahami determinan kepuasan dan ketidakpuasan pengguna aplikasi statistik pemerintah.

#### IV. SIMPULAN

Penelitian ini berhasil mengidentifikasi persepsi pengguna terhadap aplikasi AllStat BPS melalui pendekatan analisis sentimen dan pemodelan topik. Dari hasil klasifikasi sentimen menggunakan tiga algoritma, Random Forest menunjukkan kinerja terbaik dengan akurasi tertinggi dan tingkat kesepakatan prediksi yang sangat baik, sehingga dipilih sebagai model klasifikasi utama. Secara ilmiah, temuan ini memberikan bukti empiris mengenai keunggulan metode ensemble dalam konteks analisis sentimen ulasan aplikasi statistik pemerintah, yang selama ini masih terbatas pada penggunaan model klasifikasi sederhana. Selanjutnya, pendekatan Latent Dirichlet Allocation (LDA) menghasilkan delapan topik utama, yang sebagian besar mengarah pada aspek functional suitability dan performance efficiency berdasarkan kerangka ISO/IEC 25010.

Pemodelan topik yang dilakukan menunjukkan bahwa LDA efektif dalam menggali isu-isu utama dari ulasan pengguna, meskipun beberapa topik masih memiliki koherensi yang rendah. Hal ini mengindikasikan bahwa pendekatan topik modeling dapat terus ditingkatkan agar menghasilkan klasifikasi topik yang lebih representatif dan bermakna. Sebagai rencana penelitian lanjutan, disarankan untuk mengeksplorasi model topik lain seperti BERTopic atau NMF, serta mempertimbangkan integrasi dengan metode deep learning. Dengan demikian, penelitian ini tidak hanya memberikan kontribusi metodologis melalui perbandingan model analisis sentimen dan pemodelan topik, tetapi juga memperkaya literatur empiris mengenai evaluasi aplikasi statistik berbasis ulasan pengguna, serta dapat dimanfaatkan sebagai dasar evaluasi strategis dalam pengembangan aplikasi AllStat BPS di masa mendatang.

#### DAFTAR PUSTAKA

- [1] United Nations Statistics Division, "UNdata," 2024.
- [2] World Bank, *Data For Better Life*, vol. 16, no. 2. 2021.
- [3] United Nations Statistics Division, "National Reporting Platforms," 2021. <https://unstats.un.org/capacity-development/UNSD-FCDO/national-reporting-platforms/>
- [4] Badan Pusat Statistik, "Leaflet aplikasi Android Allstats [Pamflet]," no. 021, 2022.
- [5] Badan Pusat Statistik, "Analisis Hasil Survei Kebutuhan Data BPS 2022," pp. 1–141, 2022.
- [6] Badan Pusat Statistik, "Allstats BPS – Google Play Store," 2025.
- [7] H. Luthfiatun Nisa and Muhammad Hasan Sidiq Kurniawan, "Analisis Sentimen Terhadap Komentar Aplikasi Allstats BPS Dengan Klasifikasi Naïve Bayes," *Emerg. Stat. Data Sci. J.*, vol. 2, no. 3, pp. 315–329, 2024, doi: 10.20885/esds.vol2.iss.3.art24.
- [8] R. Apriani and D. Gustian, "Analisis Sentimen Dengan Naïve Bayes Terhadap Komentar Aplikasi Tokopedia," *J. Rekayasa Teknol. Nusa Putra*, vol. 6, no. 1, pp. 54–62, 2019, doi: 10.52005/rekayasa.v6i1.86.
- [9] E. Fitri, "Analisis Sentimen Terhadap Aplikasi Ruangguru Menggunakan Algoritma Naive Bayes, Random Forest Dan Support Vector Machine," *J. Transform.*, vol. 18, no. 1, pp. 71–80, 2020, doi: 10.26623/transformatika.v18i1.2317.
- [10] B. A. Maulana, M. J. Fahmi, A. M. Imran, and N. Hidayati, "Analisis Sentimen Terhadap Aplikasi Pluang Menggunakan Algoritma Naive Bayes dan Support Vector Machine (SVM)," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 2, pp. 375–384, 2024, doi: 10.57152/malcom.v4i2.1206.
- [11] J. P. Arisula and P. Parjito, "Comparison of Naive Bayes and Random Forest Methods in Sentiment Analysis on the Getcontact Application," *J. Tek. Inform.*, vol. 5, no. 5, pp. 1221–1230, 2024, doi: 10.52436/1.jutif.2024.5.5.2004.
- [12] N. L. P. Merawati, A. Z. Amrullah, and Ismarmiyati, "Analisis Sentimen dan Pemodelan Topik Pariwisata Lombok Menggunakan Algoritma Naive Bayes dan Latent Dirichlet Allocation," *J. RESTI*, vol. 5, no. 1, pp. 123–131, 2021, doi: 10.29207/resti.v5i1.2587.
- [13] Nuradilla, S., Kamila, S. A., Zahra, L., Suhaeni, C., & Sartono, B. (2025). Pemodelan Topik pada Komentar YouTube Arra: Komparasi LDA dan K-Means Menggunakan Fitur Leksikal dan Semantik. *Jurnal Informatika: Jurnal Pengembangan IT*, 10(3), 694-707.
- [14] Nikmatun, I. Alvi, Waspada, and Indra, "Implementasi Data Mining Untuk Klasifikasi Masa Studi Mahasiswa Menggunakan Algoritma K-Nearest Neighbor," *J. SIMETRIS*, vol. 10, no. 2, pp. 421–432, 2019.
- [15] F. Putra, H. F. Tahiyat, R. M. Ihsan, Rahmaddeni, and L. Efrizoni, "Application of K-Nearest Neighbor Algorithm Using Wrapper as Preprocessing for Determination of Human Weight Information," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. January, pp. 273–281, 2024, doi: 10.1109/I2CT45611.2019.9033691.
- [16] F. Koto and G. Y. Rahmamingtyas, "Inset lexicon: Evaluation of a word list for Indonesian sentiment analysis in microblogs," *Proc. 2017 Int. Conf. Asian Lang. Process. IALP 2017*, vol. 2018-January, no. December, pp. 391–394, 2017, doi: 10.1109/IALP.2017.8300625.
- [17] A. Go, R. Bhayani, and L. Huang, "Twitter Sentiment Classification using Distant Supervision," *Processing*, vol., pp. 1–6, 2009.
- [18] D. Alita and A. R. Isnain, "Pendeteksian Sarkasme pada Proses Analisis Sentimen Menggunakan Random Forest Classifier," *J. Komputasi*, vol. 8, no. 2, pp. 50–58, 2020, doi: 10.23960/komputasi.v8i2.2615.
- [19] Remawati, D., Noersasongko, E., & Marjuni, A. (2024, February). Mental Health Detection with TF-IDF Feature Extraction. In 2024

- IEEE International Conference on Artificial Intelligence and Mechatronics Systems (AIMS) (pp. 1-6). IEEE.
- [20] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "From data mining to knowledge discovery in databases," *AI Mag.*, vol. 17, no. 3, pp. 37–53, 1996.
- [21] A. Syaifuddin, R. A. Harianto, and J. Santoso, "Analisis Trending Topik untuk Percakapan Media Sosial dengan Menggunakan Topic Modelling Berbasis Algoritme LDA," *J. Intell. Syst. Comput.*, vol. 2, no. 1, pp. 12–19, 2021, doi: 10.52985/insyst.v2i1.150.
- [22] A. Tripathy, A. Agrawal, and S. K. Rath, "Classification of Sentimental Reviews Using Machine Learning Techniques," *Procedia Comput. Sci.*, vol. 57, pp. 821–829, 2015, doi: 10.1016/j.procs.2015.07.523.
- [23] H. Axelborn and J. Berggren, "Topic Modeling for Customer Insights : A Comparative Analysis of LDA and BERTopic in Categorizing Customer Calls," 2023, [Online]. Available: <https://um.kb.se/resolve?urn=urn:nbn:se:umu:diva-209225>
- [24] M. R. Fahlevi, "Sentiment Analysis And Topic Modeling on User Reviews of Online Tutoring Applications Using Support Vector Machine and Latent Dirichlet Allocation," *Knowbase Int. J. Knowl. Database*, vol. 2, no. 2, p. 142, 2022, doi: 10.30983/knowbase.v2i2.5906.
- [25] E. Suryadi, Devaldi Akbar Sulistiyani, "Evaluation of Information Quality Using ISO/IEC 25010:2011 (Case Research: Menu Harianku Application)," *Int. J. Innov. Enterp. Syst.*, vol. 6, no. 1, pp. 85–95, 2024, doi: 10.25124/ijies.v6i01.165.