

BAB I

PENDAHULUAN

1.1 Latar Belakang

Berdasarkan data *Speedtest Global Index* per Maret 2024, kecepatan internet *mobile* Indonesia untuk download sebesar 25,83 Mbps dan upload sebesar 12,54 Mbps. Sedangkan kecepatan internet *fixed broadband* Indonesia untuk download sebesar 29,37 Mbps dan upload sebesar 18,04 Mbps. Faktor-faktor yang mempengaruhi rendahnya kecepatan internet di Indonesia mencakup kapasitas jaringan, bandwidth, kualitas infrastruktur termasuk penggunaan kabel serat optik yang lama atau tidak terawat, serta penggunaan perangkat dan perangkat lunak yang ketinggalan zaman oleh pengguna. Lenny Septiani (2024).

Peningkatan proporsi waktu yang dihabiskan untuk terhubung ke layanan Wi-Fi di Indonesia, dari 33% pada kuartal pertama 2022 menjadi 37% pada kuartal ketiga 2024, menunjukkan adanya tren positif dalam akses dan pemanfaatan internet *fixed broadband*. Meskipun angka ini mungkin terkesan kecil, namun perubahan ini mencerminkan beberapa faktor penting yang sedang terjadi dalam transformasi digital di Indonesia. Berdasarkan laporan yang dibuat oleh *Opensignal*, ketergantungan terhadap layanan Wi-Fi juga sangat bervariasi di seluruh Indonesia. Di beberapa kabupaten di Indonesia menunjukkan kurang dari 30% waktu terhubung ke Wi-Fi, terutama di Sumatera, Sulawesi atau di Kepulauan Maluku. Di lain pihak, di banyak kabupaten di Jawa Timur, tercatat persentase waktu yang sangat tinggi di Wi-fi, seringkali melebihi 50%. Dengan demikian, akses ke konektivitas internet *fixed broadband* yang handal menjadi semakin penting di Indonesia. Dicky Yuniarto (2025).

Industri telekomunikasi di Indonesia telah mengalami pertumbuhan yang signifikan, seiring dengan perkembangan teknologi informasi dan komunikasi yang pesat. PT Telkom Indonesia, XL Axiata, dan First Media, sebagai pemain utama dalam industri ini, telah berperan dalam menyediakan berbagai layanan telekomunikasi, termasuk layanan *broadband* Internet melalui merek yang dikenal sebagai Indihome, XL, dan First Media. Octasylya dan Rurianto (2020).

Layanan Indihome, XL, dan First Media telah menjadi pilihan utama bagi banyak rumah tangga di Indonesia untuk memenuhi kebutuhan akses internet dan layanan telekomunikasi lainnya. Dalam menghadapi persaingan yang semakin ketat dan ekspektasi pelanggan yang semakin tinggi, pemahaman yang mendalam tentang persepsi dan sentimen pelanggan terhadap layanan ini menjadi sangat penting. Tannady dkk (2022).

Saat ini, penggunaan aplikasi My IndiHome, My XL, dan My First Media semakin meningkat karena memungkinkan pelanggan untuk mengakses berbagai layanan, melakukan pembayaran, dan melaporkan gangguan terhadap layanan Internet tersebut ketika aplikasi dapat diunduh secara gratis di App Store pada iOS dan Play Store. Tim Peneliti Puslitbang SDPPI (2018).

Sehingga, penelitian ini akan melakukan analisis sentimen pelanggan terhadap layanan Indihome, My XL, dan My First Media dengan memanfaatkan data dari ulasan dan rating di Play Store. Penelitian ini juga akan menganalisis hubungan antara rating dan ulasan pengguna, serta mengimplementasikan metode



Support Vector Machine (SVM) untuk analisis sentimen. Hasil penelitian diharapkan dapat memberikan manfaat sekaligus menambah pengetahuan untuk berbagai pihak.

1.2 Teori

1.2.1 Konsep Dasar Sistem

Sistem adalah sebuah entitas yang tidak dapat dipisahkan, terdiri dari komponen-komponen yang saling terkait atau bersambungan satu sama lain, dengan tujuan untuk mencapai suatu target tertentu. Selain itu, sistem juga dapat didefinisikan sebagai suatu pengorganisasian yang saling berinteraksi satu sama lain, saling tergantung, dan terintegrasi dalam satu komponen. Dari pengertian tersebut dapat diartikan sistem merupakan integrasi antar kelompok organisasi dan prosedur menghasilkan data dan informasi untuk kepentingan tertentu. Rahmawati dan Bachtiar (2018).

1.2.2 Sistem Informasi

Sistem informasi merupakan suatu sistem yang digunakan di dalam suatu organisasi untuk memenuhi kebutuhan pengolahan data transaksi sehari-hari, mendukung operasional, memiliki aspek manajerial, dan mendukung kegiatan strategis organisasi. Berdasarkan pengertian tersebut maka sistem informasi adalah kumpulan berbagai bagian atau komponen dari sistem yang terhubung satu sama lain untuk mendata, memproses, menyimpan dan menyalurkan informasi yang bermanfaat pada suatu organisasi. Sanjaya (2019).

1.2.3 Rating dan Sentimen Analisis

Rating dapat diartikan sebagai penilaian atau skor yang diberikan oleh pengguna terhadap suatu produk atau layanan. *Rating* sering kali dinyatakan dalam bentuk angka atau bintang, dan dapat mencerminkan tingkat kepuasan atau ketidakpuasan pengguna. *Rating* sering digunakan sebagai indikator kualitas atau performa suatu produk atau layanan. Perusahaan dan penyedia layanan dapat menggunakan *rating* untuk mengevaluasi umpan balik pelanggan dan meningkatkan kualitas produk atau layanan.

Faktor-faktor yang mempengaruhi *rating* seperti kualitas produk, layanan pelanggan, harga, dan pengalaman pengguna dapat mempengaruhi penilaian yang diberikan oleh pengguna. Studi-studi empiris mengenai faktor-faktor ini dapat memberikan wawasan tentang apa yang dapat memotivasi atau mengecewakan pengguna dalam memberikan *rating*. Fathullah et al (2020).

Sentimen merujuk pada perasaan atau opini yang diungkapkan oleh seseorang terhadap suatu hal atau kejadian. Sentimen dalam konteks analisis dapat dinyatakan sebagai positif, negatif, atau netral. Teknik analisis sentimen melibatkan penggunaan teknik-teknik komputasional untuk mengidentifikasi, mengekstraksi, dan menilai sentimen dalam teks. Metode-metode seperti *Natural Language Processing (NLP)*, *machine learning*, dan *deep learning* dapat digunakan untuk analisis sentimen.



Analisis sentimen membantu organisasi atau perusahaan memahami pandangan pengguna terhadap produk atau layanan. Dapat digunakan untuk mengukur tingkat ketidakefektifan, mendeteksi isu atau masalah potensial, dan merespons umpan balik pengguna. Skala sentimen dapat dinyatakan dalam berbagai skala, mulai dari skala bipolar (positif dan negatif) hingga skala unipolar (netral, positif, dan negatif). Penggunaan skala ini dapat memberikan pemahaman yang

lebih mendalam tentang tingkat kepuasan atau ketidakpuasan pengguna. Cahyaningtyas et al (2021).

1.2.4 Metode Support Vector Machine (SVM)

Metode *Support Vector Machine (SVM)* adalah suatu teknik dalam pembelajaran mesin yang digunakan untuk klasifikasi dan regresi. *SVM* bertujuan untuk menemukan *hyperplane* terbaik yang memisahkan dua kelas dalam ruang fitur. *Hyperplane* ini dipilih dengan memaksimalkan margin antara dua kelas. Konsep dasar *SVM* adalah *SVM* berfokus pada konsep pemisahan *hyperplane*, yang merupakan batas keputusan untuk memisahkan dua kelas dalam ruang fitur. Margin adalah jarak antara *hyperplane* dan titik-titik terdekat dari setiap kelas. *SVM* berusaha untuk memaksimalkan margin ini. *Support vectors* adalah titik-titik data yang berada paling dekat dengan *hyperplane*, dan mereka memainkan peran kunci dalam menentukan *hyperplane*. Fungsi keputusan *SVM* menggunakan fungsi *hinge loss* untuk mengukur sejauh mana sebuah titik data berada di sisi yang benar dari margin. Tujuan Optimalisasi *SVM* adalah menemukan *hyperplane* yang meminimalkan fungsi *hinge loss* dan pada saat yang sama memaksimalkan margin. *Kernel functions* memungkinkan *SVM* menangani masalah yang tidak linear dengan mentransformasikan data ke dalam dimensi yang lebih tinggi. Beberapa jenis kernel umum melibatkan kernel linear, kernel polinomial, dan *kernel radial basis function (RBF)*. Monika dan Furqon (2018).

Regularisasi *SVM* menggunakan parameter C untuk mengontrol tingkat regularisasi, yang dapat membantu mengatasi *overfitting* atau *underfitting*. Nilai parameter C yang lebih tinggi cenderung menghasilkan batas keputusan yang lebih ketat, sementara nilai yang lebih rendah dapat menghasilkan batas keputusan yang lebih fleksibel. Multi-class lassification *SVM* dapat digunakan untuk masalah klasifikasi multikelas melalui metode *one-vs-one* atau *one-vs-all*. Kombinasi hasil dalam *one-vs-one*, hasil dari semua klasifikasi dilibatkan untuk menentukan kelas akhir. Kelebihan *SVM* efektif dalam ruang fitur yang memiliki dimensi tinggi, tahan terhadap outlier karena *SVM* memperhatikan *support vectors*, metode ini dapat tahan terhadap outlier. Penerapan *SVM* dalam sentimen analisis sering digunakan dalam klasifikasi sentimen berdasarkan teks ulasan pengguna. Keunggulan *SVM* dapat mengatasi masalah klasifikasi dengan dimensi fitur yang besar dan menangani keberagaman tata bahasa dalam teks. Studi-studi terkait review literatur yang membahas penerapan *SVM* dalam berbagai konteks dan pemecahan masalah. Ritonga dan Purwaningsih (2018).

1.2.5 Term Frequency-Inverse Document Frequency (TF-IDF)

Metode *Term Frequency-Inverse Document Frequency (TF-IDF)* adalah cara untuk memberi bobot pada hubungan antara suatu kata (*term*) dan sebuah dokumen. *TF-IDF* merupakan sebuah ukuran statistik yang digunakan untuk menilai tingkat signifikansi sebuah kata dalam sebuah dokumen atau dalam sekelompok kata. Pada dokumen tunggal, setiap kalimat dianggap sebagai satu dokumen. Frekuensi kemunculan kata dalam dokumen tersebut menunjukkan tingkat signifikansi kata tersebut di dalam dokumen tersebut. Sementara frekuensi dokumen yang muncul menunjukkan seberapa umum kata tersebut. Bobot kata akan semakin sering muncul dalam suatu dokumen dan semakin kecil jika muncul dalam



suatu teknik yang menggabungkan *term frequency (TF)* dan *inverse document frequency (IDF)*. *Term Frequency* dihitung menggunakan Persamaan (1), di mana *term* muncul t_i kali term ke- i muncul dalam dokumen ke- j . Sementara itu, *Inverse Document Frequency (IDF)* merupakan logaritma dari rasio jumlah seluruh dokumen dalam

korpus terhadap jumlah dokumen yang mengandung term yang dimaksud, sebagaimana dijelaskan secara matematis dalam Persamaan (2). Nilai *TF-IDF* diperoleh dengan mengalikan kedua nilai ini, seperti dijabarkan dalam Persamaan (3).

$$tf_i = \frac{freq_i(d_j)}{\sum_{i=1}^k freq_i(d_j)} \quad (1)$$

$$idf_i = \log \frac{|D|}{|\{d:t_i \in d\}|} \quad (2)$$

$$(tf - idf)_{ij} = tf_i(d_j) \times idf_i \quad (3)$$

Maka,

$$tf_{td} idf_t = tf_{td} \times \log \left(\frac{N}{df_t} \right) \quad (4)$$

Sumber : Salton, G., & McGill, M.J. (1983)

Dimana,

- $tf * idf$ = adalah bobot total dari *term* t
- tf_{td} = adalah jumlah kemunculan term t dalam suatu dokumen
- N = adalah total dari keseluruhan dokumen
- df_t = adalah jumlah dari keseluruhan dokumen yang mengandung term

1.2.6 Analisis Ulasan Pengguna

Analisis ulasan pengguna melibatkan pemahaman konsep analisis sentimen, ekstraksi fitur, dan teknik-teknik lainnya yang digunakan untuk menggali informasi dari ulasan yang diberikan oleh pengguna. Sentimen mengacu pada perasaan atau opini yang diungkapkan oleh seseorang terhadap suatu hal atau kejadian. Analisis sentimen adalah teknik yang digunakan untuk mengekstrak dan memahami sentimen dalam teks, terutama dalam konteks ulasan pengguna. Tujuan analisis sentiment yaitu pengukuran kepuasan pengguna untuk mengidentifikasi sentimen positif, negatif, atau netral untuk mengukur tingkat kepuasan pengguna terhadap produk atau layanan. Deteksi isu atau masalah menemukan isu-isu atau masalah yang mungkin mempengaruhi pengalaman pengguna. Metode analisis sentiment meliputi *Natural Language Processing (NLP)* untuk memproses dan memahami teks manusia secara alami. *Machine Learning* dan *Deep Learning* untuk melatih model yang dapat mengklasifikasikan sentimen. Ekstraksi fitur adalah aspek-aspek tertentu dari teks yang digunakan untuk mewakili informasi. Metode ekstraksi fitur seperti *TF-IDF (Term Frequency-Inverse Document Frequency)* atau *Word Embeddings*, digunakan untuk mengidentifikasi fitur-fitur penting dalam teks. Nurian (2023).

Aspek-aspek analisis ulasan pengguna yaitu aspek produk atau layanan yang mengidentifikasi aspek-aspek spesifik yang dibahas oleh pengguna dalam ulasannya. Pengalaman



dimana pengguna mengalami produk atau layanan tersebut. Klasifikasi kategori-kategori seperti positif, negatif, atau netral. Penggunaan skala ur sejauh mana sentimen tersebut bersifat positif atau negatif. Pentingnya an bahwa sentimen dapat bervariasi tergantung pada konteksnya. Nuansa kemampuan untuk menangkap nuansa dan kompleksitas dalam ulasan k klasifikasi *Support Vector Machines (SVM)*, *Naive Bayes*, dll. Penerapan si untuk mengklasifikasikan sentimen dalam ulasan pengguna. *Deep*

Learning (Misalnya, *Recurrent Neural Networks*). Penggunaan model *deep learning* untuk analisis sentimen yang lebih kompleks. Penerapan dalam Layanan WiFi adalah spesifikasi penelitian bagaimana analisis sentimen dapat diterapkan dalam konteks layanan WiFi. Pentingnya ulasan pengguna adalah mengapa ulasan pengguna berperan penting dalam meningkatkan kualitas layanan Wi-Fi. Evaluasi kinerja meliputi metrik evaluasi penggunaan metrik evaluasi, seperti akurasi, presisi, recall, dan F1-score, untuk mengukur kinerja model analisis sentimen. Penerapan model dalam keputusan bisnis yaitu bagaimana hasil analisis sentimen dapat memberikan wawasan yang bernilai dan membantu pengambilan keputusan bisnis. Azizah dkk (2023).

1.2.7 Layanan Wi-Fi

Layanan Wi-Fi merupakan sebuah teknologi nirkabel yang memungkinkan perangkat untuk terhubung ke jaringan lokal tanpa perlu menggunakan kabel fisik. Standar IEEE 802.11, seperti 802.11b, 802.11g, 802.11n, 802.11ac, dan 802.11ax, merinci berbagai varian implementasi teknologi WiFi yang telah berkembang seiring waktu. Keberadaan berbagai frekuensi, khususnya 2,4 GHz dan 5 GHz, memberikan pilihan dalam menyusun jaringan Wi-Fi dengan mempertimbangkan kecepatan dan jangkauan yang diinginkan. Kecepatan transmisi data atau throughput adalah indikator kritis yang menggambarkan sejauh mana jaringan dapat menyediakan layanan yang responsif. Selain itu, jangkauan yang baik dan kapasitas yang memadai memastikan bahwa pengguna dapat terhubung dengan jaringan dalam berbagai lokasi dan situasi, serta dapat menangani banyak perangkat terhubung secara bersamaan. Dalam konteks kinerja jaringan Wi-Fi, konsep *Quality of Service (QoS)* menjadi penting. Parameter *QoS* seperti jitter, latensi, dan kehilangan paket memainkan peran vital dalam menentukan pengalaman pengguna. Solusi canggih seperti teknologi *Multiple Input Multiple Output (MIMO)* dan *Beamforming* membantu meningkatkan kinerja Wi-Fi dengan memperkuat sinyal dan meminimalkan interferensi. Namun, beberapa tantangan seperti interferensi dari perangkat lain di frekuensi yang sama atau keterbatasan jangkauan tetap relevan dan memerlukan penanganan khusus. Pusvita dan Huda (2019).

Manajemen lalu lintas, termasuk teknik-teknik seperti *Quality of Service (QoS)* dan load balancing, membantu memastikan alokasi sumber daya jaringan yang efisien dan memberikan prioritas pada jenis lalu lintas tertentu, seperti suara atau video, untuk mendukung aplikasi yang memerlukan kinerja *real-time*. Dalam mendesain dan mengoptimalkan layanan Wi-Fi, pemahaman mendalam terhadap preferensi dan kebutuhan pengguna menjadi kunci, dan hal ini dapat ditemukan melalui analisis ulasan pengguna. Analisis tersebut memungkinkan pemahaman tentang kepuasan pengguna, serta masalah atau tantangan yang mungkin dihadapi, sehingga memungkinkan penyedia layanan untuk merespons secara proaktif dan meningkatkan kualitas layanan Wi-Fi secara keseluruhan. Hermalia dkk (2019).

1.2.8 Hubungan Antara SVM dan Analisis Ulasan Pengguna

Integrasi *Support Vector Machine (SVM)* dalam analisis ulasan pengguna membuka peluang besar yang kompleks antara faktor-faktor yang memengaruhi penilaian dengan konsep pemisahan *hyperplane*-nya, mampu mengatasi masalah dalam ulasan pengguna. Dengan mengambil rating sebagai variabel target, pandangan yang mendalam tentang konsistensi antara *rating* numerik dan p dalam ulasan teks. Kernel functions yang digunakan oleh SVM menjadi i struktur data yang tidak linear, sehingga membantu mengekstraksi pola kompleks dalam ulasan. Penerapan *SVM* dapat memberikan pemahaman yang



lebih holistik tentang bagaimana faktor-faktor tertentu dalam ulasan pengguna dapat berdampak pada penilaian keseluruhan. Ini melibatkan identifikasi *support vectors* yang mencerminkan pola yang signifikan dalam data, dan dengan demikian, memperoleh model yang mampu menggambarkan hubungan intrinsik antara sentimen dan rating. Azizah dkk (2023).

Selain itu, *SVM* dapat membantu mengatasi masalah *overfitting* atau *underfitting* yang mungkin muncul dalam analisis ulasan pengguna. Parameter *C* dalam *SVM* dapat disesuaikan untuk mengontrol tingkat regularisasi, memastikan bahwa model tidak terlalu disesuaikan dengan data pelatihan dan dapat menggeneralisasi dengan baik pada data baru. Oleh karena itu, penerapan *SVM* bukan hanya memberikan hasil akurat dalam klasifikasi sentimen, tetapi juga memberikan tingkat interpretabilitas yang tinggi. Dalam konteks analisis konsistensi rating dan sentimen, *SVM* memungkinkan pengidentifikasian hubungan yang lebih mendalam antara ulasan dan rating yang diberikan oleh pengguna. Dengan cara ini, *SVM* berfungsi sebagai alat analisis yang kuat untuk mengeksplorasi variabilitas dalam tanggapan pengguna, sehingga membantu perusahaan atau penyedia layanan untuk meningkatkan kualitas layanan mereka berdasarkan temuan yang diperoleh dari model *SVM*. Dengan pendekatan ini, penggunaan *SVM* dapat memberikan wawasan yang kaya dan kontekstual, membuka peluang untuk pengambilan keputusan yang lebih baik dalam pengembangan dan peningkatan layanan berbasis ulasan pengguna. Fathullah dkk (2020).

1.2.9 Penelitian Terkait

Tabel 1 Penelitian Terkait

Judul	Penulis	Penerbit	Metode	Hasil
<i>Sentiment Analysis</i> menggunakan <i>Support Vector Machine (SVM)</i>	Petrix Nomleni	Jurusan Teknik Elektro Fakultas Teknologi Industri Institut Teknologi Sepuluh Nopember	<i>Support Vector Machine (SVM)</i>	Pengujian dilakukan sebanyak 7 kali dengan tingkat akurasi rata-rata di atas 80%. Akurasi tertinggi 84.4086% dengan data training 20% dan data uji 80%
Analisis Sentimen dengan <i>SVM</i> , <i>Naive Bayes</i> dan <i>KNN</i> untuk Studi	Fajar Sodik Pamungkas, Iqbal Kharisudina	Jurusan Matematika Universitas Negeri Semarang	<i>Support Vector Machine (SVM)</i> , <i>Naive Bayes</i> dan <i>K-Nearest Neighbor (KNN)</i>	Pengujian dengan metode <i>SVM</i> dengan kernel linear dan evaluasi model <i>10-fold cross validation</i> menghasilkan akurasi sebesar 90,1%.



Judul	Penulis	Penerbit	Metode	Hasil
Analisis Sentimen Wacana Pemindahan Ibu Kota Indonesia Menggunakan Algoritma <i>Support Vector Machine (SVM)</i>	Primandani Arsi, Retno Waluyo	Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)	<i>Support Vector Machine (SVM)</i>	Pengujian dengan metode <i>SVM</i> lebih baik dari metode sebelumnya (<i>BM25 + KNN dan Naive Bayes</i>). Dengan akurasi =96,68%, precision=95.82%, recall=94.04% dan AUC = 0,979.
Analisis Sentimen Gojek Pada Media Sosial Twitter Dengan Klasifikasi <i>Support Vector Machine (SVM)</i>	Nur Fitriyah, Budi Warsito, Di Asih I Maruddani	Departemen Statistika FSM Universitas Diponegoro	<i>Support Vector Machine (SVM)</i>	Hasil klasifikasi sentimen dari hasil pelabelan data secara manual menggunakan metode <i>SVM</i> pada Gojek menghasilkan tingkat akurasi keseluruhan terbaik sebesar 79,19%.
Analisis Sentimen pada <i>review</i> Aplikasi Grab di Google Play Store Menggunakan <i>Support Vector Machine</i>	Rizki Wahyudi, Gilang Kusumawar dhana	Universitas Amikom Purwokerto	<i>Support Vector Machine (SVM)</i>	Pengujian memperoleh akurasi sebesar 85,54% yang menghasilkan prediksi pada data <i>testing</i> yaitu <i>Review</i> positif sebesar 59 <i>Review</i> serta <i>Review</i> negatif sebesar 675 <i>Review</i> .

1.3 Rumusan Masalah

1. Bagaimana sentimen pelanggan yang diekspresikan melalui ulasan dan penilaian (rating)?



Metode *Support Vector Machine (SVM)* dapat diterapkan untuk menganalisis pelanggan berdasarkan data ulasan pelanggan yang telah dikumpulkan? Konsistensi antara penilaian (rating) dan ulasan berdasarkan data yang telah

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah diatas, maka penelitian ini bertujuan untuk:

1. Untuk menganalisis sentimen pelanggan terhadap layanan yang diekspresikan melalui ulasan dan penilaian (*rating*).
2. Untuk menerapkan metode *Support Vector Machine* (SVM) dalam menganalisis sentimen pelanggan berdasarkan data ulasan pelanggan yang telah dikumpulkan.
3. Untuk mengidentifikasi konsistensi antara penilaian (*rating*) pelanggan dan ulasan berdasarkan data yang telah dianalisis.

1.5 Manfaat Penelitian

Penelitian ini diharapkan memberikan manfaat sebagai berikut:

1. Pengguna, memberikan wawasan tentang sentimen pelanggan terhadap layanan yang digunakan.
2. Perusahaan, memberikan wawasan tentang area-area yang memerlukan perbaikan berdasarkan ulasan pelanggan dan dengan demikian meningkatkan kepuasan pelanggan serta menjaga daya saingnya di pasar telekomunikasi.
3. Akademisi, memberikan kontribusi pada pengetahuan dalam bidang analisis sentimen, data mining, dan penggunaan data ulasan pelanggan.

1.6 Ruang Lingkup

Penelitian ini memiliki beberapa batasan, yaitu:

1. Data yang digunakan dalam penelitian ini terbatas pada ulasan dan penilaian (*rating*) pelanggan di *platform* Play Store untuk aplikasi My IndiHome, My XL, dan My First Media.
2. Metode analisis sentimen yang digunakan dalam penelitian ini adalah metode *Support Vector Machine* (SVM).
3. Penelitian ini menggunakan data tahun 2018-2023 yang tidak mencakup seluruh sejarah aplikasi My IndiHome, My XL, dan My First Media atau perubahan signifikan dalam layanan.
4. Data yang digunakan dalam penelitian ini dalam bahasa Indonesia.
5. Penelitian ini akan memfokuskan analisis pada *rating* dan ulasan dengan skor 1 hingga 5 bintang di *platform* Play Store



BAB II

METODOLOGI PENELITIAN

2.1 Lokasi Penelitian

Penelitian ini dilakukan sejak bulan Juni 2023 di Laboratorium *Ubiquitous Computing and Networking* Teknik Informatika Universitas Hasanuddin.

2.2 Instrumen Penelitian

Instrumen yang digunakan dalam penelitian ini adalah sebagai berikut:

Tabel 2 Spesifikasi *software*

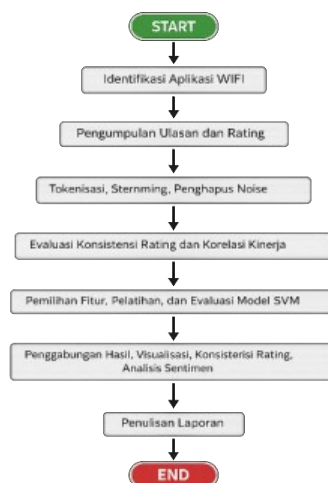
Komponen	Spesifikasi
OS	Windows 11
Bahasa	Python
Tools	Microsoft Excel, Google Collab, dan Jupyter Notebook

Tabel 3 Spesifikasi *hardware*

Komponen	Spesifikasi
Jenis	Lenovo Legion Y540
Processor	IntelCore i7
RAM	16 GB

2.3 Tahapan Penelitian

Tahapan-tahapan pada penelitian ini diperlihatkan pada Gambar 1.



Gambar 1 Tahapan Penelitian



1. Identifikasi Aplikasi Wi-Fi
Poin awal penelitian adalah mengidentifikasi aplikasi Wi-Fi yang akan dianalisis, mencakup aspek-aspek tertentu seperti fitur, keandalan, atau kinerja.
2. Pengumpulan Ulasan dan *Rating*
Mengumpulkan ulasan dan *rating* dari pengguna terkait aplikasi Wi-Fi yang menjadi fokus penelitian.
3. Tokenisasi, *Stemming*, Penghapusan *Noise*
Melakukan pre-processing terhadap data ulasan, termasuk tokenisasi untuk memecah teks menjadi kata-kata, *stemming* untuk menghasilkan bentuk kata dasar, dan penghapusan *noise* untuk membersihkan data.
4. Evaluasi Konsistensi *Rating* dan Korelasi Kinerja
Mengevaluasi konsistensi *rating* pengguna terhadap kinerja sebenarnya aplikasi Wi-Fi dan mencari korelasi antara *rating* dan kinerja.
5. Pemilihan Fitur, Pelatihan, dan Evaluasi Model *SVM*
Memilih fitur-fitur yang relevan, melatih model *SVM* untuk analisis sentimen, dan mengevaluasi kinerja model tersebut dengan menggunakan dataset pelatihan dan pengujian.
6. Penggabungan Hasil dan Visualisasi
Menggabungkan hasil analisis konsistensi *rating* dengan hasil analisis sentimen menggunakan *SVM*, serta menyajikannya secara visual, mungkin dalam bentuk grafik atau tabel.
7. Penggabungan Hasil, Visualisasi, Konsistensi *Rating*, Analisis Sentimen
Menggabungkan hasil analisis konsistensi *rating* dengan hasil analisis sentimen menggunakan *SVM*, menyajikannya secara visual, menganalisis konsistensi *rating* dan hasil analisis sentimen.
8. Penulisan Laporan
Penulis akan menyusun laporan penelitian yang akan diformat sebagai skripsi untuk mendokumentasikan proses dan hasil penelitian secara rinci.

2.4 Teknik Pengumpulan Data

- a. Studi Literatur, mengumpulkan berbagai referensi jurnal, skripsi, dan sumber informasi dari internet mengenai analisis penerapan metode *support vector machine (SVM)* pada Analisis Konsistensi *Rating* dan Sentimen Analisis Berdasarkan Ulasan Pengguna Aplikasi Layanan Wi-Fi Menggunakan Metode *SVM* serta diskusi dengan dosen pembimbing terkait judul yang diangkat.
- b. Observasi, dilakukan pada firstmedia, indihome dan myxl, dengan mengamati secara langsung alur sistem *rating* sentiment dan layanan WiFi.



data

data

akan sejumlah besar ulasan dan *rating* tentang aplikasi layanan WiFi. Data dari platform *online*, toko aplikasi, atau sumber data lainnya. Dalam tahap ini akan melibatkan seleksi dan ekstraksi data yang relevan untuk penelitian. Hasil pengumpulan dapat dilihat pada Tabel 4.

Tabel 4 Data Sampel

userName	score	at	content
iFoel NKR	3	10/09/2023 03:07	Admin... Tolong di update lh aplikasinya, tampilkan informasi paket yang aktif tiap ID, informasi kecepatan internetnya dan paket2 lainnya. agar pelanggan mudah mengetahui apa yang dia pakai. terima kasih
Alkantara Pradipta	1	10/09/2023 02:58	Sumpah nyesel ganti ke wifi ini... Wifi gak guna
Novi Adi saputra 17	1	09/09/2023 17:00	Parah sering Gangguan, Harganya Uдах Mahal,,Mending Cabut ajh ini modemnya,,Mendingan Saya Pindah ke Sebelah.
Tau Pik	1	09/09/2023 13:43	First media=First lemot. First off. First sales usill. First gangguan. Inilah provider internet dengan layanan minus.
Daniel Ardhy	1	08/09/2023 14:47	Sering sekali internat putus, tidak ada layanan internet, padahal bayarnya gak pernah nunggak, tolong diperbaiki
Calista Hardian	3	07/09/2023 04:22	WiFi sk lemot
imam nurkholis	5	07/09/2023 03:37	Jaringan baik
Nenny Handayani	5	07/09/2023 03:30	Alhamdulillah, sejak awal pakai lancar walau kadang ada gangguan tp dpt d realisasikan dng baik n cepat
IG GRAPHIES	1	06/09/2023 16:47	ping sangat tidak stabil.
Ahmad Rizqi Fadlil	1	05/09/2023 10:42	Aplikasi lemot pas buka reward nya



re

2.5.2 Preprocessing Data

Data ulasan kemudian melewati tahap preprocessing. Tahap ini mencakup beberapa proses, seperti tokenisasi (memecah teks ulasan menjadi kata-kata individual), stemming (mengubah kata menjadi bentuk dasar), serta penghapusan noise (membersihkan data dari elemen yang tidak relevan atau mengganggu). Dalam preprocessing teks Bahasa Indonesia, library Regular Expression (re), Natural Language Toolkit (NLTK), dan Sastrawi memiliki peran penting yang saling melengkapi. Library Regular Expression (re) digunakan untuk membersihkan teks dari tanda baca, angka, URL, serta karakter yang tidak diperlukan dengan memanfaatkan pola regular expression. Selanjutnya, Natural Language Toolkit (NLTK) digunakan untuk melakukan tokenisasi, yaitu memecah teks menjadi kata-kata, serta menghapus stopwords agar hanya kata-kata bermakna yang tersisa. Terakhir, Sastrawi digunakan untuk melakukan stemming, yaitu mengubah kata berimbuhan menjadi bentuk kata dasar sesuai kaidah Bahasa Indonesia, sehingga teks menjadi lebih bersih dan siap digunakan untuk analisis lanjutan. Berikut adalah tahap preprocessing data yang digunakan dalam penelitian ini:

a. *Cleaning*

Langkah pembersihan teks melibatkan identifikasi dan penghapusan karakter atau kata-kata yang tidak relevan atau tidak diinginkan. Hal ini membantu mengurangi noise dalam data dan memastikan kebersihan data. Hasil teks yang telah di bersihkan dapat dilihat pada Tabel 5.

Tabel 5 Contoh Proses *Cleaning*

<i>text_clean</i>	label
admin tolong di <i>update</i> lh aplikasinya tampilkan informasi paket yang aktif tiap id informasi kecepatan internetnya dan paket lainnya agar pelanggan mudah mengetahui apa yang dia pakaiterima kasih	<i>neutral</i>
sumpah nyesel ganti ke wifi ini wifi gak guna	<i>negative</i>
parah sering gangguanharganya udah mahal mending cabut ajh ini modernnyamendingan saya pindah ke sebelah	<i>negative</i>
<i>first media</i> <i>first</i> lemot <i>first off</i> <i>first sales</i> <i>usill</i> <i>first</i> gangguan inilah provider internet dengan layanan minus	<i>negative</i>
sering sekali internet putus tidak ada layanan internet padahal bayarnya gak pernah nunggak tolong diperbaiki	<i>negative</i>
wifi sk lemot	<i>negative</i>
	<i>positive</i>



melibatkan konversi seluruh teks ulasan ke dalam format huruf kecil. Ini menghindari perbedaan pengenalan huruf besar dan kecil yang dapat menyistensi analisis. Normalisasi teks dengan kamusalay.csv adalah proses

mengubah kata-kata tidak baku atau bahasa alay menjadi kata baku menggunakan kamus yang berisi pasangan kata alay dan padanan resminya. Proses ini dilakukan dengan membaca file *kamusalay.csv* sebagai kamus, kemudian setiap kata dalam kalimat dipisahkan (tokenisasi) dan dicek apakah terdapat dalam kamus tersebut. Jika ada, kata tersebut diganti dengan bentuk bakunya; jika tidak, dibiarkan tetap. Setelah semua kata diproses, kata-kata tersebut digabung kembali menjadi kalimat yang lebih rapi dan sesuai kaidah bahasa Indonesia. Normalisasi ini penting agar data teks menjadi lebih bersih dan konsisten sebelum digunakan dalam analisis atau pemrosesan lebih lanjut. Hasil teks yang telah di normalisasi dapat dilihat pada Tabel 6

Tabel 6 Contoh Proses Normalisasi Teks

normal	label
['admin', 'tolong', 'di', 'update', 'lh', 'aplikasinya', 'tampilkan', 'informasi', 'paket', 'yang', 'aktif', 'tiap', 'id', 'informasi', 'kecepatan', 'internetnya', 'dan', 'paket', 'lainnya', 'agar', 'pelanggan', 'mudah', 'mengetahui', 'apa', 'yang', 'dia', 'pakaiterima', 'kasih']	<i>neutral</i>
['sumpah', 'nyesel', 'ganti', 'ke', 'wifi', 'ini', 'wifi', 'tidak', 'berguna']	<i>negative</i>
['parah', 'sering', 'gangguanharganya', 'sudah', 'mahalmending', 'pergi', 'saja', 'ini', 'modemnyamendingan', 'saya', 'pindah', 'ke', 'sebelah']	<i>negative</i>
['first', 'mediafirst', 'lemot', 'first', 'off', 'first', 'sales', 'usill', 'first', 'gangguan', 'inilah', 'provider', 'internet', 'dengan', 'layanan', 'minus']	<i>negative</i>
['sering', 'sekali', 'internat', 'putus', 'tidak', 'ada', 'layanan', 'internet', 'padahal', 'bayarnya', 'tidak', 'pernah', 'nunggak', 'tolong', 'diperbaiki']	<i>negative</i>
['wifi', 'sk', 'lemot']	<i>negative</i>
['jaringan', 'baik']	<i>positive</i>
['alhamdulillah', 'sejak', 'awal', 'pakai', 'lancar', 'walau', 'kadang', 'ada', 'gangguan', 'tetapi', 'dapat', 'realisasikan', 'dng', 'baik', 'cepat']	<i>positive</i>

Sumber : Hasil Penelitian

c. Tokenisasi

Tokenisasi dilakukan untuk memecah teks ulasan menjadi unit-unit kecil, seperti kata-kata atau frasa. Hal ini membantu mengubah teks ulasan menjadi serangkaian token atau entitas yang dapat diolah secara terpisah. Hasil teks yang telah melalui proses tokenisasi dapat dilihat pada Tabel 7.



Tabel 7 Contoh Proses Tokenisasi

token	label
['admin', 'tolong', 'di', 'update', 'lh', 'aplikasinya', 'tampilkan', 'informasi', 'paket', 'yang', 'aktif', 'tiap', 'id', 'informasi', 'kecepatan', 'internetnya', 'dan', 'paket', 'lainnya', 'agar', 'pelanggan', 'mudah', 'mengetahui', 'apa', 'yang', 'dia', 'pakaiterima', 'kasih']	<i>neutral</i>
['sumpah', 'nyesel', 'ganti', 'ke', 'wifi', 'ini', 'wifi', 'tidak', 'guna']	<i>negative</i>
['parah', 'sering', 'gangguanharganya', 'sudah', 'mahalmending', 'pergi', 'saja', 'ini', 'modemnyamendingan', 'saya', 'pindah', 'ke', 'sebelah']	<i>negative</i>
['first', 'mediafirst', 'lemot', 'first', 'off', 'first', 'sales', 'usill', 'first', 'gangguan', 'inilah', 'provider', 'internet', 'dengan', 'layanan', 'minus']	<i>negative</i>
['sering', 'sekali', 'internat', 'putus', 'tidak', 'ada', 'layanan', 'internet', 'padahal', 'bayarnya', 'tidak', 'pernah', 'nunggak', 'tolong', 'diperbaiki']	<i>negative</i>
['wifi', 'sk', 'lemot']	<i>negative</i>

Sumber : Hasil Penelitian

d. Penghapusan *Stop Words*

Stop words adalah kata-kata umum yang tidak memberikan banyak informasi karena kemunculannya yang sering dalam teks. Kata-kata seperti "dan," "atau," dan "yang" dapat dihapus agar fokus lebih pada kata-kata yang lebih informatif. Hasil teks yang telah melalui proses *stop words* dapat dilihat pada Tabel 8.

Tabel 8 Contoh Proses Penghapusan *Stop Words*

stop	label
['admin', 'update', 'lh', 'aplikasinya', 'tampilkan', 'informasi', 'paket', 'aktif', 'tiap', 'id', 'informasi', 'kecepatan', 'internetnya', 'paket', 'lainnya', 'pelanggan', 'mudah', 'mengetahui', 'apa', 'pakaiterima', 'kasih']	<i>neutral</i>
['sumpah', 'nyesel', 'ganti', 'wifi', 'wifi', 'berguna']	<i>negative</i>
['parah', 'sering', 'gangguanharganya', 'mahalmending', 'pergi', 'modemnyamendingan', 'pindah', 'sebelah']	<i>negative</i>
['first', 'mediafirst', 'lemot', 'first', 'off', 'first', 'sales', 'usill', 'first', 'gangguan', 'internet', 'layanan', 'minus']	<i>negative</i>
['nat', 'putus', 'layanan', 'internet', 'padahal', 'bayarnya', 'perbaiki']	<i>negative</i>
	<i>negative</i>



['jaringan', 'baik']

*positive**Sumber : Hasil Penelitian**e. Stemming*

Proses *stemming* digunakan untuk mengubah kata-kata dalam teks ulasan menjadi bentuk kata dasarnya. Contohnya, kata-kata seperti "menggunakan," "menggunakannya," dan "menggunakanlah" akan diubah menjadi bentuk dasar "guna." Ini membantu menyederhanakan representasi kata-kata. Hasil teks yang telah melalui proses *stemming* dapat dilihat pada Tabel 8.

Tabel 9 Contoh Proses *Stopword Removal*

<i>stemmed</i>	label
['sumpah', 'nyesel', 'ganti', 'wifi', 'wifi', 'guna']	<i>negative</i>
['parah', 'sering', 'gangguanharganya', 'mahalmending', 'pergi', 'modemnyamendingan', 'pindah', 'belah']	<i>negative</i>
['first', 'mediafirst', 'lot', 'first', 'off', 'first', 'sales', 'usill', 'first', 'gangguan', 'ini', 'provider', 'internet', 'layan', 'minus']	<i>negative</i>
['sering', 'sekali', 'internat', 'putus', 'layan', 'internet', 'padahal', 'bayar', 'pernah', 'nunggak', 'baik']	<i>negative</i>
['wifi', 'sk', 'lot']	<i>negative</i>
['jaring', 'baik']	<i>positive</i>

*Sumber : Hasil Penelitian***2.5.3 Analisis Konsistensi Rating**

Evaluasi konsistensi *rating* dilakukan untuk mengukur sejauh mana peringkat yang diberikan oleh pengguna konsisten dengan kinerja sebenarnya aplikasi. Ini dapat melibatkan metode statistik untuk mengidentifikasi pola atau perbedaan signifikan dalam peringkat. Deskripsi pada *rating* dapat dilihat pada Gambar 3.

```
df_combined['score'].describe()
```



```
df_combined['score'].describe()
```

```
count    18000.000000
mean      2.428556
std       1.769232
min       1.000000
25%      1.000000
50%      1.000000
75%      5.000000
max       5.000000
```

Gambar 2 Konsistensi *Rating*

Analisis *Rating*:

- Jumlah Data (*Count*): Terdapat 18,000 data dalam sampel.
- Rata-rata (*Mean*): Rata-rata rating adalah sekitar 2.43. Ini menunjukkan bahwa rata-rata dari seluruh data adalah lebih dari 1, tetapi kurang dari 3, sehingga mayoritas rating cenderung rendah.
- Deviasi Standar (*Standard Deviation*): Deviasi standar sekitar 1.77. Ini mengukur sejauh mana data tersebar dari rata-rata. Semakin tinggi deviasi standar, semakin besar variasi dalam data. Dalam konteks ini, deviasi standar yang relatif tinggi menunjukkan bahwa rating memiliki variasi yang signifikan.
- Nilai Minimum (*Min*): *Rating* minimum adalah 1. Ini adalah *rating* terendah yang dapat diberikan.
- Kuartil (25%, 50%, 75%): Kuartil pertama (25%) dan keduanya (50%) sama, keduanya adalah 1. Ini menunjukkan bahwa 25% data pertama memiliki rating 1. Kuartil ketiga (75%) adalah 5, menunjukkan bahwa 75% data pertama memiliki rating 5 atau kurang.
- Nilai Maksimum (*Max*): *Rating* maksimum adalah 5. Ini adalah *rating* tertinggi yang dapat diberikan.

Berdasarkan analisis ini, dapat disimpulkan bahwa mayoritas data memiliki *rating* rendah (dekat dengan 1), tetapi ada variasi yang signifikan dalam rating, yang terlihat dari deviasi standar yang tinggi. Kuartil ketiga yang tinggi menunjukkan bahwa sebagian besar data berada di kisaran rating rendah, tetapi masih ada data dengan *rating* tinggi (maksimal 5).

2.5.4 TF-IDF

Pada tahap ini data yang sudah dilakukan *preprocessing* akan dihitung menggunakan TF-IDF vektor dari tiap datanya. Implementasi TF-IDF menggunakan 3 sampel data Tabel 10.



s Implementasi TF-IDF

<i>media</i>	Token	Label
/2023	['jaringan', 'baik']	Positif

03:37:16 Jaringan baik

Rizqi Fathur [‘aplikasi’, ‘tolol’, ‘muter’, ‘mulu’,
 05/09/2023 09:35:1 ‘kaga’, ‘selesai’] Negatif
 APLIKASI TOLOL MUTER2
 MULU KAGA SELESAI2

Sumber : Hasil Penelitian

Tahap *TF*, pada tahap ini setiap *term* atau fitur pada data dihitung jumlahnya pada masing masing data untuk kemudian akan dihitung pada proses selanjutnya yaitu *IDF*. Dapat dilihat pada tabel 11.

Tabel 11 Contoh Proses *TF*

Fitur	<i>TF</i>		
	D1	D2	D3
jaringan	1	0	0
baik	1	0	0
aplikasi	0	1	1
lemot	0	1	0
pas	0	1	0
buka	0	1	0
reward	0	1	0
nya	0	1	0
tolol	0	0	1
muter	0	0	1
mulu	0	0	1
kaga	0	0	1
selesai	0	0	1

Sumber : Hasil Penelitian

Berikutnya, tahap *IDF* merupakan tahap perhitungan dari *TF* yang kemudian dibagi dengan jumlah data yang ada pada data. Dapat dilihat pada persamaan 5 berikut :

$$tf_{td}idf_t = tf_{td} \times \log\left(\frac{N}{df_t}\right) \quad (5)$$

$$tf_{td}idf_t = 1 \times \log\left(\frac{3}{1}\right)$$

$$tf_{td}idf_t = 1 \times \log(3)$$



:Gill, M.J. (1983)

s *IDF*

Fitur	IDF
jaringan	0,477
baik	0,477
aplikasi	0,176
lemot	0,477
pas	0,477
buka	0,477
reward	0,477
nya	0,477
tolol	0,477
muter	0,477
mulu	0,477
kaga	0,477
selesai	0,477

Sumber : Hasil Penelitian

Maka, hasil dari *TF-IDF* pada data dapat dilihat pada Tabel 13 sebagai berikut.

Tabel 13 Contoh Hasil *TF-IDF*

Fitur	TF		
	D1	D2	D3
jaringan	0,477	0	0
baik	0,477	0	0
aplikasi	0	0,176	0,176
lemot	0	0,477	0
pas	0	0,477	0
buka	0	0,477	0
reward	0	0,477	0
nya	0	0,477	0
tolol	0	0	0,477
muter	0	0	0,477
mulu	0	0	0,477
kaga	0	0	0,477
selesai	0	0	0,477

Sumber : Hasil Penelitian

2.5.5 Analisis Sentimen Menggunakan Metode *SVM*



1, pemilihan fitur relevan dari data ulasan adalah langkah awal. Kemudian, menggunakan data latih untuk mengklasifikasikan sentimen ulasan menjadi *ral*. Evaluasi kinerja model *SVM* dilakukan dengan menggunakan dataset pernah dilihat sebelumnya.

```

import numpy as np
import matplotlib.pyplot as plt

# Dummy data, gantilah dengan data sesuai kebutuhan
labels = ['Negative', 'Neutral', 'Positive']
data1_percentages = [71.55, 14.64, 13.81]
data2_percentages = [26.43, 11.38, 62.18]
data3_percentages = [82.48, 10.93, 6.58]

x = np.arange(len(labels))
width = 0.25

fig, ax = plt.subplots(figsize=(10, 6))
colors = ['red', 'green', 'blue']

for i, (neg, neu, pos) in enumerate(zip(data1_percentages, data2_percentages,
data3_percentages)):
    rects = ax.bar(x + i * width, [neg, neu, pos], width, label=labels[i],
color=colors[i])
    for rect in rects:
        height = rect.get_height()
        ax.annotate(f'{height:.2f}%', xy=(rect.get_x() + rect.get_width() / 2, height),
            xytext=(0, 3), # 3 points vertical offset
            textcoords="offset points",
            ha='center', va='bottom')

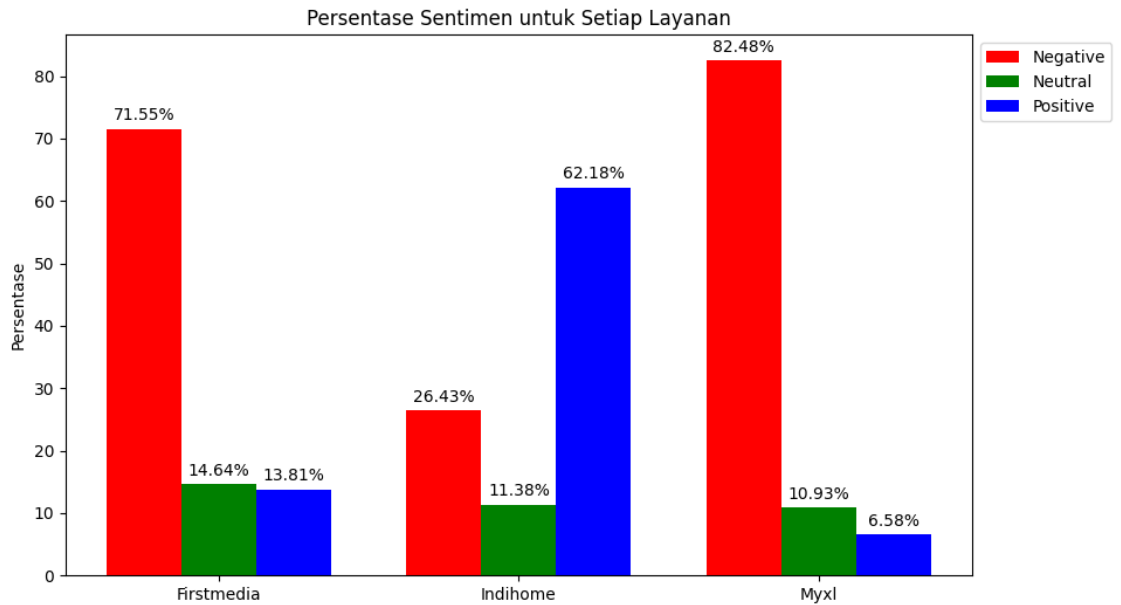
ax.set_ylabel('Persentase')
ax.set_title('Persentase Sentimen untuk Setiap Layanan')
ax.set_xticks(x + width)
ax.set_xticklabels(['Firstmedia', 'Indihome', 'Myxl'])

# Menempatkan label di sebelah kanan
ax.legend(loc='upper left', bbox_to_anchor=(1, 1))

plt.show()

```





Gambar 3 Persentase Sentimen

2.5.6 Evaluasi

Hasil dari analisis data adalah berupa evaluasi yang dilakukan dengan metode untuk perhitungan akurasi yaitu menggunakan *confusion matrix* untuk mengukur nilai *accuracy*, *precision*, dan *recall*.

