

Pengukuran Retensi Pelanggan Insyira Oleh-Oleh Berdasarkan Analisis Sentimen Pengguna Instagram

Fiki*, Inggih Permana, Febi Nur Salisah, Eki Saputra, Arif Marsal

¹Fakultas Sains dan Teknologi, Sistem Informasi, UIN Sultan Syarif Kasim Riau, Pekanbaru, Indonesia

Email: ¹11850312271@students.uin-suska.ac.id, ²inggihpermana@uin-suska.ac.id, ³febinursalisah@uin-suska.ac.id,

⁴eki.saputra@uin-suska.ac.id, ⁵arif.marsal@uin-suska.ac.id

Email Penulis Korespondensi: 11850312271@students.uin-suska.ac.id

Abstrak—Instagram sebagai salah satu platform media sosial telah membuka peluang baru bagi pelaku usaha dalam memasarkan produk secara kreatif dan efisien. Melalui fitur interaksi seperti kolom komentar, pengguna dapat menyampaikan opini terhadap produk atau layanan yang ditawarkan. Komentar-komentar ini mengandung sentimen yang dapat dijelaskan untuk memahami persepsi pelanggan. Penelitian ini bertujuan untuk mengukur retensi pelanggan menggunakan analisis sentimen pengguna Instagram. Data komentar dikumpulkan menggunakan teknik *web scraping* dari halaman Instagram, kemudian dilakukan pelabelan menggunakan leksikon dan mengklasifikasikan sentimen menjadi positif, negatif, dan netral menggunakan pendekatan analisis sentimen. Analisis ini dihubungkan dengan konsep retensi pelanggan, yang menjadi strategi penting dalam menjaga hubungan jangka panjang dengan konsumen. Retensi pelanggan dapat dilihat dari perilaku pembelian berulang, sikap loyal yang mencakup emosi, keyakinan, dan keinginan membeli kembali. Selanjutnya, hasil pengukuran retensi pada penelitian ini adalah, sentimen positif memiliki tingkat retensi sebesar 53,4% (303 dari 567 komentar), netral sebesar 6,9% (45 dari 650 komentar), dan negatif sebesar 15,1% (22 dari 146 komentar). Secara keseluruhan, total komentar yang tergolong dalam retensi mencapai 370 dari 1.363 komentar atau sebesar 27,1%. Jika dilihat dari proporsi kontribusi sentimen terhadap total retensi, komentar positif mendominasi dengan 81,9% (303 dari 370 komentar retensi). Temuan ini menunjukkan bahwa meskipun jumlah komentar netral paling banyak, komentar positif justru memberikan kontribusi terbesar terhadap retensi pelanggan. Hal ini mengindikasikan bahwa sentimen positif menjadi indikator kuat dalam mempertahankan pelanggan, sehingga perusahaan perlu mendorong terciptanya pengalaman yang positif melalui produk, pelayanan, dan interaksi aktif di media sosial.

Kata Kunci: Analisis Sentimen, Instagram; Leksikon; Retensi Pelanggan; Scraping Data

Abstract—Instagram as a social media platform has opened new opportunities for businesses to market their products creatively and efficiently. Through interactive features such as the comments section, users can express their opinions about the products or services offered. These comments contain sentiments that can be analyzed to understand customer perceptions. This study aims to measure customer retention using sentiment analysis of Instagram user comments. The comment data was collected using web scraping techniques from the Instagram page, followed by labeling using a lexicon-based approach and sentiment classification into positive, negative, and neutral categories through sentiment analysis. This analysis is linked to the concept of customer retention, which is an important strategy for maintaining long-term relationships with consumers. Furthermore, the results of customer retention analysis in this study show that positive sentiment has a retention rate of 53.4% (303 out of 567 comments), neutral sentiment 6.9% (45 out of 650 comments), and negative sentiment 15.1% (22 out of 146 comments). Overall, 370 out of 1,363 comments, or 27.1%, were categorized as contributing to retention. In terms of the proportion of sentiment contributing to total retention, positive comments dominate with 81.9% (303 out of 370). These findings suggest that although neutral comments are the most frequent, positive sentiment contributes the most to customer retention. This indicates that positive sentiment is a strong predictor of customer loyalty, highlighting the importance for companies to foster positive experiences through quality products, reliable services, and active engagement on social media. Insyira is capable of maintaining customer retention, especially from those who express positive sentiment, which reflects satisfaction with its products, services, and interactions on social media.

Keywords: Customer Retention; Data Scraping; Instagram; Lexicon; Sentiment Analysis

1. PENDAHULUAN

Sejalan dengan berkembangnya teknologi informasi, Media sosial merupakan bagian penting yang menyatu dalam kehidupan masyarakat dan berperan dalam membentuk pola komunikasi dan interaksi antar pribadi maupun kelompok. Dalam dunia bisnis, media sosial digunakan sebagai sarana promosi dan untuk berinteraksi secara langsung dengan pelanggan, membangun kedekatan, memperkenalkan merek, mengurangi biaya operasional, serta menerima masukan dengan cepat. Pelanggan dapat memberikan komentar, bertanya, dan berbagi pengalaman menggunakan produk atau layanan[1][2]. Promosi sering dianggap sama pentingnya dengan produk itu sendiri yang berfungsi untuk meningkatkan penjualan[3]. Keberadaan platform media sosial, khususnya Instagram telah membuka peluang baru bagi pelaku usaha untuk mengembangkan bisnis mereka secara kreatif dan efektif. Fitur-fitur yang membantu pelaku usaha besar maupun kecil untuk ikut memasarkan produknya dengan mudah di Instagram[4]. Instagram termasuk dalam kategori media sosial yang paling banyak diminati dan memiliki basis pengguna yang sangat banyak. Banyaknya interaksi antar pengguna melalui komentar menyebabkan terciptanya berbagai pendapat. Platform ini memberikan fasilitas bagi pengguna untuk berbagi konten berupa foto dan video, sekaligus berkomunikasi satu sama lain melalui kolom komentar. Komentar tersebut mengandung berbagai opini yang dapat di analisis melalui teknik analisis sentimen[5].

Salah satu pelaku usaha yang aktif memanfaatkan media sosial Instagram sebagai sarana promosi sekaligus penyampaian informasi produk adalah insyira oleh-oleh pekanbaru. Insyira hanya melakukan analisis pengukuran yang dilakukan terhadap *followers* mereka selama ini hanya didasarkan pada metrik *insight* dan jangkauan. Namun, Insyira



belum pernah melakukan analisis terhadap komentar yang terdapat pada akun Instagram. Pada akun Instagram insyira memiliki banyak komentar positif maupun komentar negatif, yang berfungsi sebagai masukan penting bagi perusahaan dalam mengevaluasi serta meningkatkan kualitas pelayanannya untuk meningkatkan kualitas pelayanan[6]. Pelaku bisnis dituntut untuk mampu mengelola informasi yang dimiliki, memahami pola dan kejadian masa lalu, serta melakukan prediksi masa depan secara akurat guna meningkatkan retensi pelanggan melalui strategi yang tepat dan responsif terhadap kebutuhan pasar[7]. Penilaian terhadap komentar dilakukan melalui analisis sentimen dengan mengumpulkan data dari halaman *web* baik secara manual maupun program yang disebut dengan teknik *web scraping*. Analisis sentimen merupakan proses klasifikasi perasaan dalam tulisan pada media sosial guna memperoleh informasi tentang sentimen positif, negatif, maupun netral[8].

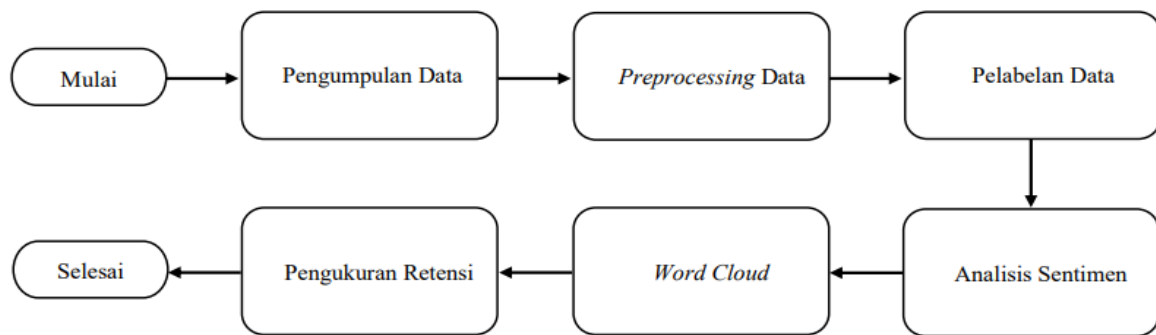
Pendekatan berbasis leksikon merupakan metode analisis sentimen yang menggunakan daftar kata opini untuk menentukan polaritas teks. Dalam tahap analisis sentimen, pendekatan ini digunakan untuk melakukan pelabelan polaritas secara otomatis positif, negatif, atau netral terhadap teks, berdasarkan kecocokan kata dalam leksikon yang telah disusun. Terdapat dua jenis pendekatan, yaitu berbasis kamus, yang memperluas leksikon melalui sinonim dan antonim dari sumber seperti WordNet, serta berbasis korpus, yang menyesuaikan kata berdasarkan konteks dalam data teks[9]. Proses ini diawali dengan penyusunan manual kata-kata opini dan dikembangkan secara bertahap guna meningkatkan akurasi analisis[10]. Penelitian ini bertujuan untuk mengukur retensi pelanggan melalui pendekatan analisis sentimen pada komentar pengguna di akun Instagram @insyraoleholehpekanbaru. Penelitian ini dijalankan dengan menggunakan bahasa pemrograman *Python* yang dioperasikan melalui perangkat lunak *PyCharm*. Metode yang diterapkan mengintegrasikan pemrosesan komputasi dan strategi retensi untuk menelusuri hubungan antara sentimen dalam komentar di platform Instagram dan retensi pelanggan. Penelitian ini memanfaatkan dataset berupa komentar dari Instagram. Komentar-komentar tersebut terlebih dahulu melalui tahap preprocessing dan pelabelan sebelum dianalisis untuk mengukur tingkat retensi. Retensi diukur berdasarkan klasifikasi sentimen yang masing-masing dikategorikan sebagai retensi atau non-retensi. Selanjutnya, data dianalisis menggunakan metode statistik seperti *Chi-Square Goodness of Fit*, *Test of Independence*, dan *Odds Ratio*. Pengujian dilakukan untuk memperoleh hasil yang komprehensif.

Retensi pelanggan merupakan strategi perusahaan dalam menjaga hubungan berkelanjutan dengan pelanggan yang telah menjadi bagian dari basis konsumen untuk meningkatkan loyalitas dan keberlanjutan bisnis[11]. Penelitian yang telah dilaksanakan oleh Qhairani, dkk (2024) memaparkan implementasi sentimen analisis pada media sosial Instagram dengan hasil mengindikasikan kepuasan pelanggan terhadap perusahaan dalam aspek produk, layanan, serta tindakan yang diambil oleh perusahaan sehingga perusahaan dianggap mampu mempertahankan pelanggannya[12]. Sibarani, dkk (2021) membahas peran data analitik dalam perumusan strategi pemasaran guna mempertahankan loyalitas pelanggan. Hasilnya menunjukkan bahwa data analitik menjadi komponen penting dalam merancang strategi pemasaran yang efisien dan berkelanjutan, yang akhirnya memberikan dampak positif terhadap tingkat retensi pelanggan[13]. Usman Hadi Yulianto (2023) membahas peran Instagram sebagai alat promosi dengan hasil optimalisasi Instagram sebagai media sosial CRM mampu menciptakan dampak positif terhadap jumlah kunjungan wisata di Jawa Timur Park 3, menunjukkan potensi besar platform ini sebagai alat pemasaran digital[14]. Sinambela dkk (2022) Mengkaji pengaruh kualitas layanan dan citra merek terhadap retensi pelanggan, dengan hasil menunjukkan bahwa keduanya memiliki dampak signifikan dalam menjaga keberlanjutan basis pelanggan [15]. Desak Gede Eka Tirtawati (2021) membahas keterkaitan antara kualitas pelayanan, kepuasan pelanggan dan retensi pelanggan dengan hasil analisis mengkonfirmasi adanya pengaruh positif kualitas pelayanan terhadap tingkat kepuasan konsumen, yang akhirnya berdampak signifikan pada peningkatan retensi nasabah di Bank BCA[16].

Penelitian ini dengan beberapa studi terdahulu memiliki kesamaan pada fokus utama yang sama, yaitu memahami dan meningkatkan retensi pelanggan melalui berbagai pendekatan seperti kualitas layanan, strategi pemasaran, kepuasan pelanggan, dan pemanfaatan media sosial. Sebagaimana penelitian sebelumnya, studi ini juga menggunakan media sosial Instagram sebagai sumber data utama untuk mengevaluasi perilaku pelanggan. Namun, penelitian ini memiliki perbedaan signifikan dalam hal pendekatan dan metode analisis yang digunakan. Penelitian ini mengintegrasikan analisis sentimen dengan menggunakan bahasa pemrograman *Python* melalui platform *PyCharm*, yang diterapkan secara spesifik pada komentar pengguna Instagram. Selain itu, pengukuran tingkat retensi pelanggan dilakukan berdasarkan klasifikasi sentimen (positif, netral, negatif) yang kemudian dianalisis lebih lanjut menggunakan metode statistik seperti *Chi-Square Goodness of Fit*, *Test of Independence*, dan *Odds Ratio*. Hasil dari penelitian ini diharapkan dapat memberikan kontribusi dalam memperluas pemahaman mengenai pemanfaatan media sosial dalam penerapan *Customer Relationship Management* (CRM), khususnya dalam upaya meningkatkan retensi pelanggan. Dengan menggunakan media sosial sebagai sumber data, penelitian ini menunjukkan bagaimana interaksi pelanggan dapat dianalisis untuk mendorong minat berkelanjutan dalam melakukan pembelian. Pemilik usaha dapat memanfaatkan temuan ini untuk mengoptimalkan sumber daya yang dimiliki dalam rangka mempertahankan pelanggan serta memberikan informasi yang relevan guna memperkuat retensi pelanggan.

2. METODOLOGI PENELITIAN

Penelitian ini menggunakan metodologi yang disusun secara sistematis untuk memastikan tercapainya tujuan penelitian. Alur tahapan penelitian secara lengkap dapat diamati pada Gambar 1.



Gambar 1. Tahapan Penelitian

2.1 Pengumpulan Data

Pengumpulan data yang digunakan berasal dari komentar pada postingan Instagram @insyiraoleholehpekanbaru. Pengumpulan data dilakukan dengan teknik *scraping* menggunakan bahasa pemrograman Python untuk mengambil komentar secara otomatis dari akun Instagram tersebut. Data komentar yang digunakan dalam penelitian ini diperoleh dari tanggal 23 mei 2024 sampai 24 mei 2025. Selama satu tahun tersebut, terkumpul sebanyak 1.363 komentar dari berbagai postingan pada akun Instagram Insyira Oleh-oleh pekanbaru.

2.2 Preprocessing Data

Pada tahap ini, data diproses melalui serangkaian tahapan prapengolahan yang meliputi pembersihan data, *case folding*, tokenisasi, dan *stemming*. Prosedur ini dilakukan untuk meningkatkan kualitas data, serta mengoptimalkan efisiensi dataset untuk tahap pengolahan selanjutnya[5]. *Case folding* merupakan tahap awal dalam praproses teks di mana seluruh huruf dalam suatu kalimat diubah menjadi huruf kecil. Karakter selain huruf abjad tidak dipertimbangkan dalam proses ini. *Tokenizing* merupakan proses pemisahan setiap komentar menjadi potongan-potongan kata (*token*) agar data siap untuk dianalisis pada tahap selanjutnya. *Stemming* adalah proses menghilangkan afiksasi seperti awalan, sisipan, akhiran, maupun gabungan dari imbuhan tersebut, sehingga kata dikembalikan ke bentuk dasarnya.

2.3 Pelabelan Data

Dalam penelitian ini, dilakukan proses pelabelan untuk mengklasifikasikan komentar-komentar di Instagram ke dalam kategori sentimen positif, netral, atau negatif[17]. Pada tahap pelabelan data, pendekatan berbasis leksikon digunakan untuk menetapkan polaritas suatu teks, yaitu menentukan apakah teks tersebut bersifat positif, negatif, atau netral, berdasarkan kecocokan kata-kata dalam leksikon yang telah disusun sebelumnya. Tahapan pelabelan ini bertujuan agar sasaran analisis dapat tercapai secara tepat sebelum memasuki tahap pengolahan data lebih lanjut. Langkah ini penting untuk memastikan bahwa data yang digunakan dalam analisis memiliki klasifikasi sentimen yang jelas dan valid, sehingga dapat menunjang ketepatan hasil penelitian.

2.4 Analisis Sentimen

Analisis sentimen adalah pendekatan otomatis yang dimanfaatkan untuk memproses data berbasis teks dan opini pengguna yang berisi pendapat, perasaan, penilaian, evaluasi, emosi, serta sikap individu terhadap produk[18]. Analisis sentimen adalah proses penambangan data untuk mengidentifikasi perasaan seseorang berdasarkan tulisan atau teks, dengan tujuan untuk klasifikasi data agar dapat memperoleh informasi mengenai sentimen positif, negatif, maupun netral yang terdapat pada media sosial[19][20]. Analisis sentimen dibagi menjadi tiga kategori, yakni sentimen positif, sentimen negatif, dan sentimen netral. Untuk menghitung persentase sentimen perkategori digunakan rumus berikut.

$$\text{Persentase sentimen perKategori} = \left(\frac{\text{jumlah komentar sentimen}}{\text{jumlah total komentar}} \right) \times 100 \quad (1)$$

2.5 Uji Analisis Sentimen

Pada tahap ini, dilakukan tiga jenis pengujian statistik untuk memastikan validitas dan kelayakan hasil penelitian. Pertama, digunakan uji *Chi-Square Goodness of Fit* untuk mengevaluasi apakah distribusi sentimen (positif, netral, negatif) dalam data bersifat seimbang atau menunjukkan kecenderungan tertentu. Kedua, dilakukan uji *Chi-Square Test of Independence* guna menilai apakah terdapat hubungan signifikan antara jenis sentimen komentar dengan tingkat retensi pelanggan. Ketiga, digunakan pengujian *Odds Ratio* sebagai analisis pelengkap untuk mengukur besar peluang retensi berdasarkan kategori sentimen. Ketiga pengujian ini digunakan untuk memperkuat temuan secara statistik.

2.5.1 Chi Square Goodness of Fit

Chi Square Goodness of Fit salah satu jenis uji *Chi-Square* (χ^2) yang digunakan untuk mengevaluasi apakah distribusi frekuensi dari data hasil observasi berbeda secara signifikan dibandingkan dengan distribusi yang diharapkan. Uji ini

berfungsi untuk menilai kesesuaian antara data nyata yang diperoleh dengan distribusi teoretis atau ekspektasi yang telah ditentukan sebelumnya[21]. Perhitungan dalam pengujian *Chi Square Goodness of Fit* mengacu pada rumus berikut.

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} \quad (2)$$

O_i merupakan frekuensi observasi (*observed frequency*) ke- i , E_i adalah frekuensi harapan (*expected frequency*) ke- i , dan simbol \sum menunjukkan penjumlahan untuk semua kategori atau kelas data.

2.5.2 Chi Square Test of Independence

Uji *Chi-Square* untuk Independensi digunakan untuk menentukan apakah terdapat hubungan atau ketergantungan antara dua variabel kategorikal, atau apakah kedua variabel tersebut bersifat independen[21]. Uji ini dapat digunakan untuk mengetahui apakah terdapat hubungan antara kategori sentimen komentar (positif, netral, negatif) dengan status retensi pelanggan. Proses pengujian ini diawali dengan menyusun tabel kontingensi, yang dapat dilihat pada Tabel 1.

Tabel 1. Tabel Kontingensi

| Sentimen | Retensi iya | Retensi tidak | Total baris (i) |
|-----------------|-------------|---------------|-----------------|
| Positif | Jumlah | Jumlah | Total jumlah |
| Negatif | Jumlah | Jumlah | Total jumlah |
| Netral | Jumlah | Jumlah | Total jumlah |
| Total kolom (j) | Total iya | Total tidak | Jumlah total |

Tabel kontingensi ini menggambarkan hubungan antara kategori sentimen (positif, negatif, netral) dan status retensi pelanggan (iya atau tidak). Setiap sel menunjukkan jumlah komentar berdasarkan kombinasi sentimen dan retensi. Total baris menunjukkan jumlah komentar per sentimen, sedangkan total kolom menunjukkan jumlah per kategori retensi. Jumlah total mencakup seluruh data komentar yang dianalisis. Tabel ini berguna untuk melihat pola hubungan antara sentimen pelanggan dan tingkat retensi. Dalam melakukan perhitungan frekuensi harapan untuk *chi square test of independence* digunakan rumus berikut.

$$E_{ij} = \frac{\text{total baris}_i \times \text{total kolom}_j}{\text{total keseluruhan}} \quad (3)$$

E_{ij} merupakan frekuensi harapan pada sel baris ke- i dan kolom ke- j . Total baris- i adalah jumlah total pada baris ke- i , Total kolom- j merupakan jumlah total pada kolom ke- j , sedangkan Total keseluruhan adalah jumlah dari seluruh data yang tersedia.

2.5.3 Odds Ratio

Odds Ratio merupakan suatu ukuran statistik yang digunakan untuk menggambarkan kekuatan hubungan serta perbandingan peluang antara dua kelompok. *Odds ratio* bertujuan untuk melengkapi hasil dari uji *Chi-Square* yang telah dilakukan, dengan memberikan ukuran efek yang menjelaskan ada atau tidaknya keterkaitan antar kelompok, serta memungkinkan interpretasi hasil dalam bahasa yang lebih sederhana. Untuk melakukan penghitungan *Odds Ratio* digunakan rumus berikut.

$$OR = \frac{a \times d}{b \times c} \quad (4)$$

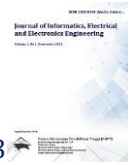
Variabel a menunjukkan jumlah kejadian positif pada kelompok pertama, b adalah jumlah kejadian negatif pada kelompok pertama, c merupakan jumlah kejadian positif pada kelompok kedua, dan d adalah jumlah kejadian negatif pada kelompok kedua.

2.6 Word Cloud

Word cloud merupakan pengembangan dari situs jejaring sosial berbasis *web*. Hasil word cloud menyajikan representasi visual dari data ulasan pelanggan, kata-kata yang memiliki frekuensi tinggi dalam review ditampilkan dengan ukuran huruf yang lebih besar. Semakin banyak suatu kata muncul dalam sebuah teks, maka semakin besar pula representasinya dalam visualisasi gambar. Istilah ini juga dikenal sebagai *tag cloud*, yaitu representasi visual yang menampilkan frekuensi kemunculan kata dari suatu materi tertulis yang dipilih, teks dalam buku, atau konten dari situs *web*[22].

2.7 Pengukuran Retensi Pelanggan

Retensi pelanggan adalah strategi yang difokuskan pada upaya perusahaan dalam mempertahankan konsumen yang telah di miliki agar terus melakukan interaksi bisnis secara berkelanjutan [11][23]. Retensi pelanggan berhubungan dengan perilaku loyal yang ditunjukkan melalui intensitas pembelian terhadap produk tertentu. Loyalitas tersebut tidak hanya bersifat perilaku, tetapi juga dipengaruhi oleh sikap seperti keyakinan, emosi, dan keinginan untuk membeli kembali[24]. Terdapat 4 dimensi retensi pelanggan yaitu pembelian berulang, ukuran pesanan rata-rata, pembelian beberapa produk, rujukan ke pembeli lain di luar tempat usaha[25]. Merancang sebuah modul berbasis pola (rule-based) yang berfungsi untuk mendeteksi sinyal retensi dalam komentar. Modul ini dibangun dengan memanfaatkan ekspresi reguler (*regular expressions*), yakni metode pencocokan pola teks yang umum digunakan dalam proses pengolahan bahasa alami (*Natural*



Language Processing). Untuk menghitung persentase retensi komentar label yang telah dilakukan dengan modul pola, yang dirancang menggunakan *RegEx* dapat menggunakan rumus berikut.

$$\text{Retensi Komentar label} = \left(\frac{\text{jumlah retensi dari komentar label}}{\text{jumlah komentar label}} \right) \times 100 \quad (5)$$

Untuk menghitung retensi secara keseluruhan dapat menggunakan rumus berikut.

$$\text{Retensi Total} = \left(\frac{\text{jumlah total retensi}}{\text{jumlah total komentar}} \right) \times 100 \quad (6)$$

3. HASIL DAN PEMBAHASAN

Sumber data penelitian diperoleh dari tanggapan atau komentar pengguna pada unggahan di platform Instagram Insyira Pekanbaru. Pengumpulan data dilakukan dengan teknik *scraping* menggunakan bahasa pemrograman Python untuk mengambil komentar secara otomatis dari akun Instagram tersebut. Data komentar yang digunakan dalam penelitian ini diperoleh dari tanggal 23 mei 2024 sampai 24 mei 2025. Selama satu tahun tersebut, terkumpul sebanyak 1.363 komentar dari berbagai postingan pada akun Instagram Insyira Oleh-oleh pekanbaru.

3.1 Preprocessing Data

Pada tahap ini, *preprocessing* mencakup pembersihan data dengan menghapus semua *hashtag*, *mention*, dan emotikon yang ada dalam komentar. Selain itu, pada tahap *preprocessing* ini juga dilakukan *case folding* proses mengubah suatu kalimat menjadi huruf kecil. Selanjutnya, *tokenizing* proses pemisahan setiap komentar menjadi potongan-potongan kata, dan *stemming* yakni proses menghilangkan afiksasi seperti awalan, sisipan, akhiran, maupun imbuhan, sehingga kata dikembalikan ke bentuk dasarnya. Tahapan *preprocessing* data ini dapat diamati pada Tabel 2.

Tabel 2. Preprocessing Data

| Tahapan | Hasil |
|---------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Komentar Awal | Wahh makin nyaman ni belanja di insyira kalo udah pake AC, langganan tahun ke tahun kalo mau pulang kampung pasti borong oleh oleh nya di insyira. Emang ga mau berpaling ke toko lain serius |
| Case Folding | wah makin nyaman ni belanja di insyira kalau udah pakai ac langganan tahun ke tahun kalau mau pulang kampung pasti borong oleh oleh nya di insyira memang tidak mau berpaling ke toko lain serius |
| Tokenizing | ['wah', 'makin', 'nyaman', 'ni', 'belanja', 'di', 'insyira', 'kalau', 'udah', 'pakai', 'ac', 'langganan', 'tahun', 'ke', 'tahun', 'kalau', 'mau', 'pulang', 'kampung', 'pasti', 'borong', 'oleh', 'oleh', 'nya', 'di', 'insyira', 'memang', 'tidak', 'mau', 'berpaling', 'ke', 'toko', 'lain', 'serius'] |
| Stemming | ['semakin', 'nyaman', 'belanja', 'di', 'insyira', 'kalau', 'sudah', 'pakai', 'ac', 'langgan', 'tahun', 'ke', 'tahun', 'kalau', 'mau', 'pulang', 'kampung', 'pasti', 'borong', 'oleh', 'oleh', 'di', 'insyira', 'memang', 'tidak', 'mau', 'paling', 'ke', 'toko', 'lain', 'serius'] |

3.2 Pelabelan

Pelabelan dilakukan agar target analisis dapat tercapai secara tepat sebelum dilakukan tahap pengolahan data lebih lanjut. proses pelabelan dengan pendekatan berbasis leksikon dilakukan dengan memberikan kategori pada komentar-komentar Instagram dengan memberikan nilai positif, netral, atau negatif. Adapun potongan hasil pelabelan komentar telah diklasifikasikan berdasarkan jenis sentimen, sebagaimana ditunjukkan pada Tabel 3.

Tabel 3. Pelabelan

| No | Komentar | Sentimen |
|------|-------------------------------------------------------------------------------------------------------------------------------------|----------|
| 1 | recommended buat oleh oleh sawah lunto | Positif |
| 2 | tata kelola parkir lalu macet parkir depan toko ganggu arus lalu lintas macet tidak berkah banyak orang marah tiap lewat depan toko | Negatif |
| 3 | pisang kipas frozen berapa harga masing masing ukur boleh tahu | Netral |
| ... | ... | ... |
| 1425 | sukses selalu buat insyira oleh oleh pekanbaru semoga semakin laris langgan | Positif |

Komentar positif menunjukkan kepuasan, apresiasi, dan loyalitas pelanggan, serta mencerminkan pengalaman menyenangkan, kepercayaan terhadap kualitas, dan rekomendasi yang memperkuat citra perusahaan. Komentar netral bersifat informatif tanpa emosi yang jelas, biasanya berupa pertanyaan atau permintaan informasi dari pelanggan yang belum memiliki pengalaman langsung. Komentar negatif berisi ketidakpuasan, keluhan, atau kritik, yang menunjukkan aspek layanan yang perlu diperbaiki oleh Insyira. Setelah proses pelabelan dilakukan, data komentar kemudian dianalisis menggunakan bahasa pemrograman Python untuk mengidentifikasi komentar yang mengindikasikan retensi pelanggan.

Proses ini melibatkan pengelompokan berdasarkan hasil analisis sentimen. Berikut potongan hasil komentar yang masuk kategori retensi dapat dilihat pada Tabel 4.

Tabel 4. Komentar yang Terindikasi Retensi

| Username | Komentar | Retensi |
|-------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------|
| @ devi_umma_arkhano | kalau ke pekanbaru pasti ke insyira selalu ramai tidak pernah sepi | Retensi |
| @ linda_putri_aqila | toko oleh selalu kalau mau balik kampung selalu belanja ke sini | Retensi |
| @suci.ramaaadona | sebelum viral bisa beli banyak santai | Retensi |
| @denisha_ayuni | bus wisata edukasi sekolah sman kerinci kanan siak setiap tahun study tour | Retensi |
| @m_neliyanti | pasti singgah beli oleh di insyira masya allah tabarakallah insyira pusat oleh oleh lengkap murah di pekanbaru | Retensi |
| @ dela_tristiansi_putri | selalu di rekomendasi buat tamu tamu pusat semakin nyaman belanja di insyira kalau sudah pakai ac langgan tahun ke tahun kalau mau pulang kampung pasti borong oleh oleh di insyira memang tidak mau paling ke toko lain serius | Retensi |

Tabel 4 menampilkan komentar-komentar yang terindikasi retensi, yaitu komentar dari pelanggan yang menunjukkan kepuasan terhadap produk, menyatakan telah melakukan pembelian dalam jumlah banyak, berencana untuk kembali membeli, merekomendasikan produk kepada orang lain, atau secara rutin berlangganan. Indikasi ini mencerminkan adanya potensi retensi pelanggan yang kuat. Dalam konteks ini, terdapat empat dimensi utama retensi pelanggan yang digunakan sebagai acuan, yaitu: pembelian berulang, ukuran pesanan rata-rata, pembelian berbagai jenis produk, serta rujukan atau rekomendasi kepada pembeli lain di luar tempat usaha.

3.3 Hasil Analisis Sentimen

Berdasarkan analisis statistik terhadap 1.363 komentar pelanggan, distribusi sentimen menunjukkan bahwa sebagian besar komentar bersifat netral, yaitu sebanyak 650 komentar atau 47,7%, diikuti oleh sentimen positif sebanyak 567 komentar atau 41,6%, dan negatif sebanyak 146 komentar atau 10,7%. Komentar dengan sentimen netral umumnya berisi pertanyaan atau informasi seputar produk, sementara komentar positif banyak mengandung apresiasi terhadap pelayanan dan kualitas produk yang diberikan oleh Insyira Oleh-Oleh Pekanbaru. Secara keseluruhan, dominasi sentimen netral dan positif dalam komentar pelanggan menunjukkan bahwa Insyira Oleh-Oleh Pekanbaru telah berhasil membangun interaksi yang cukup baik di media sosial, khususnya melalui platform Instagram. Hal ini mencerminkan tingkat kepuasan pelanggan terhadap kualitas produk, responsivitas layanan, serta pendekatan komunikasi yang dilakukan oleh perusahaan. Meskipun komentar negatif jumlahnya relatif kecil, keberadaannya tetap penting sebagai indikator area yang perlu dievaluasi, seperti antri, kemacetan, ketersediaan stok produk, atau respons layanan. Analisis sentimen ini dapat menjadi landasan untuk memperkuat hubungan dengan pelanggan, serta meningkatkan retensi dan loyalitas di masa mendatang. Dengan memahami pola dan persepsi yang tercermin dalam komentar, perusahaan memiliki peluang untuk lebih adaptif dalam menyesuaikan pelayanan sesuai harapan konsumen. Adapun hasil analisis sentimen yang diperoleh dari data komentar Instagram, mencakup distribusi sentimen, tingkat retensi tiap kategori, serta karakteristik utama masing-masing sentimen, dapat dilihat pada Tabel 5.

Tabel 5. Hasil Analisis Sentimen

| Sentimen | Tingkat Retensi | Karakteristik Kunci |
|----------|-----------------|---------------------------------------------|
| Positif | 41,6% | Pujian, kepuasan, rekomendasi ke orang lain |
| Netral | 47,7% | Pertanyaan informatif mengenai produk |
| Negatif | 10,7% | Keluhan, kritik, dan ketidakpuasan |

3.4 Uji Analisis Sentimen

Pada tahap ini dilakukan tiga jenis pengujian guna memastikan validitas dan kelayakan dari penelitian. Pengujian pertama adalah uji *chi-square goodness of fit*, dilanjutkan dengan uji *chi-square test of independence* untuk mengukur hubungan antara sentimen dan retensi pelanggan, serta diakhiri dengan uji *odds ratio* sebagai pengujian tambahan.

3.4.1 Uji Chi Square Goodness of Fit

Hasil uji *Chi-Square Goodness of Fit* menunjukkan adanya perbedaan yang sangat signifikan antara distribusi sentimen aktual dengan distribusi yang diharapkan merata ($\chi^2 = 452.15$, $p < 0.0001$). Dari total 1.363 komentar pelanggan, distribusi aktual menunjukkan pola yang tidak merata, di mana sentimen netral mendominasi sebanyak 650 komentar (47,7%), diikuti oleh sentimen positif sebanyak 567 komentar (41,6%), dan negatif sebanyak 146 komentar (10,7%). Jika distribusi sentimen diasumsikan merata, maka frekuensi harapan untuk masing-masing kategori adalah 454,33 komentar. Nilai *Chi-Square* yang tinggi (452.15) menunjukkan adanya perbedaan substansial antara frekuensi aktual dan yang diharapkan. *P-value* yang sangat kecil ($p < 0.0001$) menguatkan bahwa perbedaan ini tidak disebabkan oleh faktor kebetulan, melainkan mencerminkan pola nyata dalam data. Ukuran efek *Cramer's V* sebesar 0.408 menunjukkan adanya ketidakseimbangan yang signifikan dengan kekuatan hubungan yang tergolong moderat. Dominasi sentimen netral dan positif

Copyright © 2025 The Author, Page 170



This Journal is licensed under a Creative Commons Attribution 4.0 International License



REFERENCES

- [1] A. A. P. Moh. Rizki Fahreza, "Crm Di Era Media Sosial: Memanfaatkan Platform Sosial Untuk Meningkatkan Interaksi Pelanggan," *J. Manajemen, Akuntansi, Dan Ekon.*, Vol. 6, No. 1, Pp. 61–70, 2024, Doi: <https://doi.org/10.8734/Musyari.V6i1.4034>.
- [2] R. R. Agam Arsyad, Yaqub Dalimunthe, "Penerapan Sosial Media Dalam Strategi Meningkatkan Umkm," *Neraca Manajemen, Ekon.*, Vol. 3, No. 1, Pp. 111–121, 2024, Doi: <https://doi.org/10.8734/Musyari.V3i6.1803>.
- [3] M. D. Yuniar, "Aplikasi Instagram Sebagai Sarana Promosi Bisnis Online : Sebuah Studi Pada Akun Warung Mamak," *Intelektiva J. Ekon. Sos. Hum.*, Vol. 2, No. 3, Pp. 1–9, 2020, [Online]. Available: <https://www.jurnalintelektiva.com/index.php/jurnal/article/view/298>
- [4] S. A. Quraniah And I. Mayasari, "Analisis Penerapan Customer Relationship Management Pada Website Dan Media Sosial Instagram (Studi Kasus Pada Shopee, Toko Kopi Tuku, Zalora, Dan Pt. Tirta Investama (Aqua)," *J. Manaj. Dan Bisnis Madani*, Vol. 3, No. 1, Pp. 1–19, 2022, Doi: 10.51353/Jmbm.V3i1.530.
- [5] A. Tazkia And Y. Arkhiansyah, "Implementasi Naive Bayes Classifier Dalam Menganalisis Sentimen Pelanggan Mie Gacoan Pada Instagram," *J. Tek.*, Vol. 7, No.3, No. X, Pp. 139–149, 2023, Doi: 10.5281/Zenodo.10851075.
- [6] M. W. A. Putra, Susanti, Erlin, And Herwin, "Analisis Sentimen Dompot Elektronik Pada Twitter Menggunakan Metode Naive Bayes Classifier," *It J. Res. Dev.*, Vol. 5, No. 1, Pp. 72–86, 2020, Doi: 10.25299/Itjrd.2020.Vol5(1).5159.
- [7] N. Susi, S. Sugiana, And B. Musty, "Analisis Data Sistem Informasi Monitoring Marketing; Tools Pengambilan Keputusan Strategic," *Jutisi J. Ilm. Tek. Inform. Dan Sist. Inf.*, Vol. 12, No. 2, Pp. 696–708, 2023, [Online]. Available: <https://repository.lpkia.ac.id/id/eprint/9/1/1240-3383-1-Pb.Pdf>
- [8] H. Syah And A. Witanti, "Analisis Sentimen Masyarakat Terhadap Vaksinasi Covid-19 Pada Media Sosial Twitter Menggunakan Algoritma Support Vector Machine (Svm)," *J. Sist. Inf. Dan Inform.*, Vol. 5, No. 1, Pp. 59–67, 2022, Doi: 10.47080/Simika.V5i1.1411.
- [9] S. A. Bahrainian And A. Dengel, "Sentiment Analysis Using Sentiment Features," *Proc. - 2013 Ieee/Wic/Acm Int. Jt. Conf. Web Intell. Intell. Agent Technol. - Work. Wi-Iatw 2013*, Vol. 3, No. 13, Pp. 26–29, 2013, Doi: 10.1109/Wi-Iatw.2013.145.
- [10] G. Gautam And D. Yadav, "Sentiment Analysis Of Twitter Data Using Machine Learning Approaches And Semantic Analysis," *2014 7th Int. Conf. Contemp. Comput. Ic3 2014*, Pp. 437–442, 2014, Doi: 10.1109/Ic3.2014.6897213.
- [11] D. S. Wahyuni, E. Sunarya, And N. Norisanti, "Analisis Retensi Pelanggan Dan Switching Costs Terhadap Loyalitas Pelanggan Indihome," *J. Manag. Bussines*, Vol. 4, No. 1, Pp. 99–107, 2022, Doi: 10.31539/Jomb.V4i1.741.
- [12] Q. F. F. Safiesza, M. Afdal, And R. Novita, "Sentimen Analisis Social Crm Pada Media Sosial Instagram Menggunakan Machine Learning Untuk Mengukur Retensi Pelanggan," *Build. Informatics, Technol. Sci.*, Vol. 6, No. 2, Pp. 606–616, 2024, Doi: 10.47065/Bits.V6i2.5269.
- [13] B. Erik Sibarani, S. Setiawan, T. Hadi, T. Williams, And T. Mkhize, "Penggunaan Data Analistik Dalam Strategi Pemasaran Untuk Mempertahankan Loyalitas Pelanggan," Vol. 3, No. 1, Pp. 30–39, 2024, Doi: 10.33050/Mentari.V3i1.592.
- [14] U. H. Yulianto, "Dampak Penerapan Sosial Crm Dan Peran Instagram Sebagai Alat Promosi Pariwisata (Jawa Timur Park3)," *Jesya*, Vol. 6, No. 2, Pp. 2230–2241, 2023, Doi: 10.36778/Jesya.V6i2.1215.
- [15] E. A. Sinambela, E. Retnowati, E. Ernawati, U. P. Lestari, And M. Munir, "Pengaruh Kualitas Layanan Dan Citra Perusahaan Terhadap Retensi Pelanggan Bengkel Resmi Honda Surabaya," *J. Baruna Horiz.*, Vol. 5, No. 1, Pp. 17–25, 2022, Doi: 10.52310/Jbhorizon.V5i1.73.
- [16] D. G. Eka Tirtawati, "Keterkaitan Antara Kualitas Pelayanan, Kepuasan Pelanggan Dan Retensi Pelanggan Di Sektor Perbankan," *J. Entrep. Manag. Ind.*, Vol. 4, No. 2, Pp. 49–55, 2021, Doi: 10.36782/Jemi.V4i2.2168.
- [17] N. M. S. Hadna, I. S. Paulus, And W. Winarno, "Studi Literatur Tentang Perbandingan Metode Untuk Proses Analisis Sentimen Di Twitter," *Semin. Nas. Teknol. Inf. Dan Komun.*, No. March, Pp. 1–8, 2016, [Online]. Available: <https://id.scribd.com/document/540279328/95>
- [18] A. Faesal, A. Muslim, A. H. Ruger, And K. Kusri, "Sentimen Analisis Terhadap Komentar Konsumen Terhadap Produk Penjualan Toko Online Menggunakan Metode K-Means," *Matrik J. Manajemen, Tek. Inform. Dan Rekayasa Komput.*, Vol. 19, No. 2, Pp. 207–213, 2020, Doi: 10.30812/Matrik.V19i2.640.
- [19] D. Safryda Putri And T. Ridwan, "Analisis Sentimen Ulasan Aplikasi Pospay Dengan Algoritma Support Vector Machine," *J. Ilm. Inform.*, Vol. 11, No. 01, Pp. 32–40, 2023, Doi: 10.33884/Jif.V11i01.6611.
- [20] N. A. Rakhmawati, M. I. Aditama, R. I. Pratama, And K. H. U. Wiwaha, "Analisis Klasifikasi Sentimen Pengguna Media Sosial Twitter Terhadap Pengadaan Vaksin Covid-19," *J. Inf. Eng. Educ. Technol.*, Vol. 4, No. 2, Pp. 90–92, 2020, Doi: 10.26740/Jieet.V4n2.P90-92.
- [21] Y. Cahyono And S. Saprudin, "Analisis Sentiment Tweets Berbahasa Sunda Menggunakan Naive Bayes Classifier Dengan Seleksi Feature Chi Squared Statistic," *J. Inform. Univ. Pamulang*, Vol. 4, No. 3, Pp. 87–94, 2019, Doi: 10.32493/Informatika.V4i3.3186.
- [22] F. Miley And A. Read, "Using Word Clouds To Develop Proactive Learners," *J. Scholarsh. Teach. Learn.*, Vol. 11, No. 2, Pp. 91–110, 2011, [Online]. Available: <http://search.ebscohost.com/login.aspx?direct=true&db=eric&an=Ej932148&site=ehost-live>
- [23] F. Sudirjo, T. Purwati, W. Widyastuti, Y. U. Budiman, And ..., "Analisis Dampak Strategi Pemasaran Digital Dalam Meningkatkan Loyalitas Pelanggan: Perspektif Industri E-Commerce," *J. Pendidik. ...*, Vol. 7, Pp. 7524–7532, 2023, [Online]. Available: <https://jptam.org/index.php/jptam/article/view/7422%0ahttps://jptam.org/index.php/jptam/article/download/7422/6136>
- [24] Y. Ibrahim, T. Abbas, And M. Kamal, "The Impact Of Online Communities-Based Social Customer Relationship Management (S-Crm) On Customer Loyalty And Brand Image On Hotels," *J. Assoc. Arab Univ. Tour. Hosp.*, Vol. 0, No. 0, Pp. 0–0, 2021, Doi: 10.21608/Jaauth.2021.84867.1202.
- [25] M. Edward And S. Sahadev, "Role Of Switching Costs In The Service Quality, Perceived Value, Customer Satisfaction And Customer Retention Linkage," *Asia Pacific J. Mark. Logist.*, Vol. 23, No. 3, Pp. 327–345, 2011, Doi: 10.1108/13555851111143240