



# Pengukuran Akuisisi Pelanggan Insyira Oleh-Oleh Berdasarkan Analisis Sentimen Pengguna Instagram

M. Roid Wira Mulia\*, Inggih Permana, Febi Nur Salisah, Eki Saputra, Arif Marsal

Fakultas Sains dan Teknologi, Sistem Informasi, Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Indonesia

Jl. HR. Soebrantas, 28293, Pekanbaru, Indonesia

Email: <sup>1</sup>11850312390@students.uin-suska.ac.id, <sup>2</sup>inggihpermana@uin-suska.ac.id, <sup>3</sup>febinursalisah@uin-suska.ac.id,

<sup>4</sup>eki.saputra@uin-suska.ac.id, <sup>5</sup>arif.marsal@uin-suska.ac.id

Email Penulis Korespondensi: 11850312390@students.uin-suska.ac.id

**Abstrak**—Media sosial, terutama Instagram, telah mengubah cara bisnis berinteraksi dengan pelanggan dan memasarkan produk. Namun, masih terdapat gap dalam literatur tentang pengukuran akuisisi pelanggan melalui analisis sentimen komentar di Instagram. Penelitian ini bertujuan untuk mengukur akuisisi pelanggan di Insyira Oleh-Oleh Pekanbaru dengan menganalisis 1.363 komentar dari Mei 2024 hingga Mei 2025 menggunakan *Natural Language Processing* (NLP) berbasis *Python*. Hasil menunjukkan sentimen netral mendominasi (47.7%) dengan tingkat akuisisi tertinggi (50.9%) atau setiap 2 komentar netral 1 akuisisi dibandingkan komentar positif (37.7%) dan negatif (41.8%). Uji *Chi-square* menegaskan signifikansi hubungan antara sentimen dan akuisisi, dengan nilai  $\chi^2=21.78$ ;  $p<0.0001$  dan (OR=0.58; CI[0.46,0.73]) yang mengindikasikan peluang akuisisi positif 42% lebih rendah dibanding komentar netral dan membentuk konsistensi triangular yang mengeliminasi keraguan. Sentimen negatif juga memberikan akuisisi yang lebih tinggi dibanding sentimen positif sehingga temuan ini menantang asumsi bahwa komentar positif paling efektif dalam akuisisi, menunjukkan bahwa komentar netral yang berisi pertanyaan produk memiliki potensi akuisisi lebih besar. Penelitian ini memberikan wawasan baru dalam strategi pemasaran digital, menekankan pentingnya respons cepat terhadap komentar netral untuk meningkatkan konversi pelanggan baru.

**Keywords:** Akuisisi Pelanggan; Analisis Sentimen; *Chi-square*; Instagram; *Natural Language Processing*; *Odds Ratio*.

**Abstract**—Social media, especially Instagram, has transformed how businesses interact with customers and market products. However, there remains a literature gap regarding customer acquisition measurement through sentiment analysis of Instagram comments. This research aims to measure customer acquisition at Insyira Oleh-Oleh Pekanbaru by analyzing 1,363 comments from May 2024 to May 2025 using Python-based Natural Language Processing (NLP). The results show neutral sentiment dominates (47.7%) with the highest acquisition rate (50.9%) - meaning every 2 neutral comments yield 1 acquisition - compared to positive (37.7%) and negative comments (41.8%). The Chi-square test confirms the significant relationship between sentiment and acquisition ( $\chi^2=21.78$ ;  $p<0.0001$ ), while (OR=0.58; CI[0.46,0.73]) indicates positive comments have 42% lower acquisition probability than neutral ones, forming triangular consistency that eliminates doubts. Negative sentiment also yields higher acquisition than positive sentiment, challenging the assumption that positive comments are most effective for acquisition. This reveals neutral comments containing product inquiries have greater acquisition potential. The study provides new insights for digital marketing strategy, emphasizing the importance of quick responses to neutral comments to enhance new customer conversion.

**Keywords:** Chi-square; Customer Acquisition; Instagram; Natural Language Processing; Odds Ratio; Sentiment Analysis.

## 1. PENDAHULUAN

Media sosial telah menjadi bagian yang tidak terpisahkan dari kehidupan masyarakat modern. Media sosial tidak hanya digunakan untuk berkomunikasi, tetapi juga sebagai alat yang sangat efektif dalam membangun merek, memasarkan produk, dan menjangkau audiens secara. Media sosial telah mengubah pola interaksi digital secara fundamental, bergeser dari komunikasi satu-arah menjadi model banyak pihak [1]. Instagram, sebagai platform media sosial, telah berkembang menjadi alat yang sangat kuat dalam strategi digital marketing. Dengan fitur-fitur seperti Instagram *Stories*, *Reels*, dan iklan berbayar, platform ini memungkinkan bisnis untuk menjangkau audiens yang lebih luas dengan cara yang kreatif dan menarik. Platform Instagram yang diproyeksikan memiliki 103 juta pengguna di Indonesia pada 2025 mendorong transformasi pemasaran digital [2][3]. Usaha Mikro Kecil dan Menengah (UMKM) merupakan tulang punggung ekonomi Indonesia yang berkontribusi signifikan pada peningkatan kesejahteraan masyarakat berpendapatan rendah sekaligus pelestarian tradisi dan budaya lokal [4].

Sebagai contoh, UMKM Insyira (@insyiraoleholehpekanbaru), yang berlokasi di Jalan Arifin Ahmad, Pekanbaru, memanfaatkan Instagram untuk promosi produk dan interaksi pelanggan sejak 2020. Dengan 136.000 pengikut dan 15.545 postingan (per Mei 2025), Insyira dipilih sebagai studi kasus karena berhasil memanfaatkan media sosial untuk menjangkau pelanggan dari berbagai daerah. Pemilihan Insyira juga didasarkan pada tingginya interaksi pengguna di akun Instagram mereka, terlihat dari traffic aktif dan banyaknya komentar di setiap unggahan. Ini menunjukkan tingkat engagement yang kuat, menjadikannya contoh relevan untuk mengevaluasi efektivitas strategi akuisisi pelanggan berbasis media sosial. Bagi UMKM seperti Insyira, Instagram menjadi nadi pertumbuhan bisnis, namun tanpa metrik terukur, interaksi belum tentu berujung pada transaksi. Pemasaran konten strategis di media sosial terbukti meningkatkan brand awareness dan konversi penjualan melalui pendekatan tidak langsung [5]. Akuisisi pelanggan adalah proses strategis dimana perusahaan berusaha mendapatkan pelanggan baru yang sebelumnya belum pernah membeli atau menggunakan



produk mereka. Proses ini mencakup mengubah prospek menjadi pelanggan aktual. Dengan demikian, akuisisi pelanggan mencerminkan keberhasilan perusahaan dalam mempengaruhi keputusan konsumen hingga terjadi transaksi [6]. Akuisisi pelanggan merujuk pada proses identifikasi dan perolehan pelanggan baru dan menguntungkan untuk memperluas pangsa pasar serta meningkatkan basis pelanggan perusahaan [7]. Strategi akuisisi dapat dioptimalkan melalui pemodelan klasifikasi yang membantu menentukan segmen pasar tepat, memprediksi perilaku pelanggan, dan merancang kampanye terarah. Fokus utamanya adalah menarik pelanggan baru, baik dari pasar belum terjamah maupun kompetitor dengan memanfaatkan prediksi berbasis data guna meningkatkan efektivitas dan efisiensi strategi. Pendekatan ini memungkinkan perusahaan memfokuskan sumber daya pada kontak paling relevan sekaligus membangun hubungan saling menguntungkan jangka panjang. *Customer Relationship Management* (CRM) menjadi hal penting pengelolaan relasi pelanggan jangka panjang [8], dioptimalkan melalui teknik akuisisi berbasis data seperti *web scraping* untuk ekstraksi data masif [9] dan analisis sentimen guna memahami motivasi konsumen [10]. Analisis Sentimen adalah proses sistematis untuk mengekstrak, mengolah, dan mengkategorikan opini dari data teks guna memahami sikap/persepsi publik secara komputasional, oleh karena itu hasilnya digunakan untuk pengambilan keputusan bisnis, riset pasar, atau prediksi tren [11].

Penelitian terdahulu, seperti analisis sentimen, social CRM dan akurasi pengujian dataset berdasarkan komentar di sosial media telah dilakukan. Namun, belum ada penelitian yang mengukur akuisisi pelanggan berdasarkan sentimen analisis pengguna instagram. Penelitian-penelitian tersebut belum menjawab pertanyaan krusial: Bagaimana hubungan antara pola sentimen komentar dan tingkat akuisisi aktual, Apakah komentar positif benar-benar paling efektif menarik pelanggan baru. Penelitian ini mencari apakah ada hubungan antara analisis sentimen bahasa indonesia dengan pengukuran akuisisi pelanggan dari komentar instagram, khususnya akun instagram @insyiraoleholehekanbaru. Penelitian menggunakan bahasa *python*, dan aplikasi *PyCharm*. *Python* juga sangat efektif untuk pengembangan komersial, *web scraping*, serta penelitian ilmiah basis komputasi [12]. Penelitian ini mengintegrasikan pendekatan komputasi dan strategi akuisisi pelanggan untuk mengidentifikasi hubungan antara sentimen komentar pada media sosial Instagram dan tingkat akuisisi pelanggan. Dalam penelitian ini, implementasi CRM berbasis *Natural Language Processing* (NLP) menghadapi tantangan kompleksitas data tidak terstruktur dan skalabilitas [13].

Penelitian ini menggunakan dataset komentar instagram, kemudian komentar akan dilakukan *Preprocessing* dan pelabelan data yang akan diproses untuk dilakukan pengukuran akuisisi. Data yang diukur dalam akuisisi berdasarkan positif, negatif dan netral ini, memiliki akuisisi atau non-akuisisi yang nantinya akan di uji dengan metrik *Chi-square Goodness of Fit*, *Test of Independence* dan *Odds Ratio*. Data diuji dengan komprehensif. Pada Penelitian yang dilakukan oleh Mahsa Heidari dan Pirooz shamsinejad dalam pembuatan dataset instagram untuk analisis sentimen bahasa persia dengan metode *crowdsourcing* mendapatkan 110,000 komentar instagram dari *web scraping* [14] dan pada Iqbal Darmawan dan Indriati untuk analisis sentimen terhadap akuisisi saham tokopedia oleh tiktok menggunakan *naïve bayes* berdasarkan komentar youtube menghasilkan rata tertinggi sebesar 80,514% pada perlakuan mempertahankan negasi. *Confusion matrix* akurasi 80,51%, presisi 82,54%, *recall* 77,40%, dan *F1-score* 79,89% [15]. Dan Penelitian yang dilakukan Qhairani Frilla F. Safiesza tentang analisis sentimen pada social CRM di Instagram menggunakan machine learning menunjukkan bahwa algoritma SVM dan *Naïve Bayes* mencapai akurasi tertinggi sebesar 74,26%, sedangkan *Random Forest* 63,97%. Analisis terhadap 1.945 komentar (531 positif, 868 netral, 546 negatif) menunjukkan bahwa pelanggan cenderung bertahan pada perusahaan [16].

Pada Penelitian Adhithia Erfina dkk. menggunakan *Naïve Bayes Classifier* untuk analisis sentimen terhadap vaksin COVID-19 pada 100 tweet. Untuk kata kunci "covid vaccine", diperoleh sentimen positif 93%, negatif 72%, netral 35%, dengan akurasi 94,74%. Sedangkan untuk "corona vaccine", akurasi mencapai 75,47% [17]. Penelitian Levina Marthina Rumaropen mengenai efektivitas strategi *Customer Relationship Management* PT POS Indonesia menunjukkan bahwa Instagram terbukti menjadi platform CRM lebih efektif, dengan *Fitur Stories* dan *Direct Message* meningkatkan keterlibatan dan kepuasan pelanggan [18]. Penelitian Heidari dan Safiesza telah membangun dasar kuat untuk analisis sentimen Instagram, namun belum menyentuh aspek akuisisi pelanggan UMKM [14][16]. Darmawan mengukur akuisisi saham secara bisnis tapi di platform non-Instagram [15]. Penelitian Adhithia menunjukkan akurasi yang tinggi dengan didominasi sentimen positif, Namun tidak mengeksplorasi ketertarikan penduduk untuk vaksin [17]. Penelitian Levina menegaskan efektivitas instagram dalam CRM [18]. Berdasarkan penelitian-penelitian sebelumnya, penulis menggabungkan beberapa komponen, yaitu pengukuran akuisisi pelanggan berdasarkan analisis sentimen pengguna Instagram, dengan mengintegrasikan temuan-temuan terdahulu ke dalam penelitian ini.

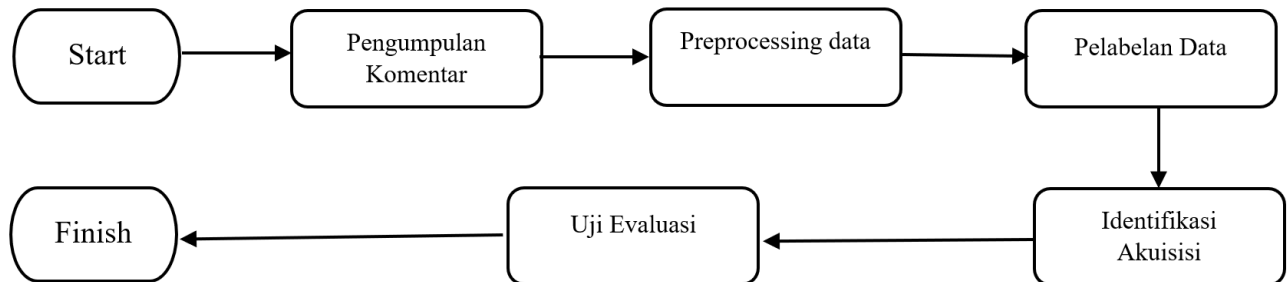
Penelitian ini bertujuan untuk memperkenalkan pendekatan baru dalam mengukur akuisisi pelanggan berbasis analisis sentimen dari komentar pengguna Instagram, khususnya pada akun @insyiraoleholehekanbaru. Dengan mengintegrasikan analisis sentimen berbasis *Natural Language Processing* (NLP) dan pengukuran akuisisi pelanggan menggunakan metrik statistik, penelitian ini memberikan kontribusi baru dalam literatur pemasaran digital. Diharapkan hasil penelitian ini dapat memberikan wawasan terhadap penggunaan media sosial dalam penerapan *Customer Relationship Management* (CRM) dan meningkatkan ketertarikan pelanggan untuk datang dan membeli produk. Hasil Penelitian ini diharapkan dapat memberikan wawasan terhadap penggunaan sosial media dalam penerapan CRM terutama akuisisi untuk pengambilan data dan meningkatkan ketertarikan pelanggan untuk datang dan membeli. Pemilik usaha dapat mengoptimalkan seluruh sumber daya yang tersedia untuk perluasan usaha dan memberikan informasi lebih kepada

calon pelanggan. Penelitian ini diharapkan menjadi landasan untuk penelitian selanjutnya dalam data yang lebih inklusif dan mengembangkan penemuan akuisisi berkelanjutan yang akan membuka literatur baru dalam pemasaran digital.

## 2. METODOLOGI PENELITIAN

### 2.1 Tahapan Penelitian

Untuk metodologi penelitian terdapat 5 tahap yang dilakukan. Berikut tahapan penelitian yang dilakukan dapat dilihat pada Gambar 1



Gambar 1. Tahapan Penelitian

Gambar 1 menunjukkan tahapan penelitian yang dimulai dari pengumpulan komentar, dilanjutkan dengan preprocessing data, pelabelan data, dan uji evaluasi. Hasil evaluasi digunakan untuk mengidentifikasi akuisisi pelanggan berdasarkan analisis sentimen.

### 2.2 Pengumpulan Komentar

Pengumpulan data dilakukan dengan studi literatur tentang text mining, yang merupakan metode analisis data yang digunakan untuk mengekstraksi informasi penting dari teks untuk mengidentifikasi dan menganalisis pola, tren, dan topik yang relevan dengan topik tertentu. Ini dapat dilakukan dengan menggunakan analisis sentimen untuk mengidentifikasi sentimen yang terdapat pada pernyataan [19]. Selanjutnya melakukan observasi ke akun Instagram untuk melihat postingan yang tersedia di akun Instagram untuk dilakukan scraping. *Web scraping* merupakan suatu teknik yang digunakan untuk mendapatkan suatu data atau informasi pada website tertentu. Informasi yang didapat berupa, teks, tautan, video, audio ataupun dokumen. Teknik ini dilakukan dengan menggunakan Bahasa pemrograman *Python* dan *Beautifulsoup4 library* dan *extension browser Brave* yaitu *Instant Data Scraper*. Penelitian Teknik *Web Scraping* dilakukan dengan batasan-batasan yang ditetapkan [20].

### 2.3 Preprocessing Data

Preprocessing adalah suatu proses untuk menyeleksi data teks agar menjadi lebih terstruktur dengan melalui serangkaian tahapan, hasil dari text *preprocessing* adalah daftar-daftar kata yang mempunyai arti yang penting yang dapat memberikan arti terhadap konten yang dokumen atau kumpulan dari dokumen [21]. 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*, *tokenizing*, penghapusan *stopword*, dan *stemming*. Dapat dilihat sebagai berikut:

- Case Folding* Mengonversi seluruh karakter dalam teks ke huruf kecil untuk menyeragamkan format.
- Tokenizing*. Pada tahap ini, penulis memecah kalimat menjadi unit kata atau frasa individual guna memfasilitasi analisis lebih lanjut.
- Stopword*. Pada tahap ini, penulis Menyaring kata-kata penghubung, atau kata ganti yang tidak relevan (misalnya: "dan", "di", "saya").
- Stemming*. Pada tahap ini, penulis mengurangi kata berimbuhan (awalan, sisipan, akhiran) ke bentuk dasar (contoh: "berjalan" → "jalan").

### 2.4 Pelabelan Data

Pelabelan data dilakukan untuk menentukan kategori yang didapatkan oleh Analisis sentiment, analisis sentimen diberikan oleh pengguna internet pada media sosial untuk memberikan suatu penilaian atau opini pribadi [22]. Dengan *lexicon* berperan sebagai basis data kata/frasa berlabel sentimen (misal: "puas"=positif, "kecewa"=negative, "bertanya"=netral) yang menjadi pondasi klasifikasi [11]. Analisis sentimen juga sebuah proses menganalisis teks untuk menentukan sentimen yang diungkapkan, seperti positif, negatif, atau netral, dapat digunakan untuk berbagai tujuan, seperti memahami opini publik, mengukur kepuasan pelanggan [23]. Pendekatan *lexicon rule-based* mengombinasikan *lexicon* ini dengan aturan linguistik untuk menentukan polaritas teks secara otomatis [24]. *Lexicon* disini dilakukan



dengan kombinasi dari kompilasi *InsetLexicon* [25] dan dikurasi secara manual. Hasil dari tahap ini akan menghasilkan tiga kategori data, yaitu komentar positif, netral, dan negatif.:

- Komentar Positif Merujuk pada ulasan yang mengandung ekspresi kepuasan, apresiasi, saran konstruktif, atau rekomendasi *Lexicon* rule-based mengidentifikasi kata kunci positif dan imbuhan kata.
- Komentar Netral. Meliputi pertanyaan atau permintaan informasi tambahan terkait produk/jasa tanpa menyertakan penilaian emosional. Aturan *lexicon* mendeteksi tidak adanya kata berpolaritas
- Komentar negatif. Mencakup kritik, keluhan, atau ketidakpuasan pelanggan terhadap kualitas layanan atau produk. *Lexicon* mengenali kata negatif dan rule-based menangkap kata negasi atau tolakan.

## 2.5 Identifikasi Akuisisi

Pada tahapan ini, dilakukan identifikasi akuisisi berdasarkan tahapan sebelumnya yaitu, analisis sentiment dan pelabelan data. Untuk melakukan identifikasi ini, penulis mengembangkan sebuah modul berbasis pola (*rule-based*) yang mampu mengenali sinyal akuisisi pada komentar Modul ini dirancang menggunakan ekspresi reguler (*regular expressions*), yaitu teknik pencocokan pola teks yang lazim digunakan dalam pemrosesan bahasa alami [26], untuk mengenali: Pola minat beli: "beli", "harga", "chat", "stok" (skor +1.0 per pola). Pola ketertarikan: "enak", "mantap", "recommended" (skor +0.5 per pola). Setiap komentar dianalisis dan diberi skor berdasarkan jumlah kecocokan terhadap pola-pola tersebut. Jika total skor memenuhi ambang batas yaitu  $\geq 0.5$ , maka komentar tersebut diklasifikasikan berpotensi sebagai akuisisi Perhitungan persentase sentimen menggunakan rumus berikut:

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

Sentimen disini ada 3, yaitu positif, negative, dan netral. Untuk contoh sentimen positif bisa dilihat seperti berikut:

$$\text{Persentase sentimen Positif} = \left( \frac{\text{jumlah komentar Positif}}{\text{jumlah total komentar}} \right) \times 100 \quad (2)$$

Untuk menghitung persentase akuisisi yang telah dilakukan dengan modul pola, yang dirancang menggunakan *RegEx* itu seperti ini:

$$\text{Akuisisi Komentarlabel} = \left( \frac{\text{jumlah akuisisi dari komentarlabel}}{\text{jumlah komentarlabel}} \right) \times 100 \quad (3)$$

Yang dihitung menjadi label positif :

$$\text{Akuisisi Komentar Positif} = \left( \frac{\text{jumlah akuisisi dari komentar positif}}{\text{jumlah komentar positif}} \right) \times 100 \quad (4)$$

Dan untuk total Akuisisi keseluruhan yaitu seperti :

$$\text{Akuisisi Total} = \left( \frac{\text{jumlah total akuisisi}}{\text{jumlah total komentar}} \right) \times 100 \quad (5)$$

## 2.6 Uji Evaluasi

Dalam uji evaluasi ini, penulis mengambil 3 pengujian yaitu uji *Chi-square goodness of fit*, *Chi-square test of independence*, dan uji *odds ratios*.

### 2.6.1 Chi-square Goodness of Fit

*Chi-square Goodness of Fit* salah satu jenis uji statistik untuk mengevaluasi apakah suatu model teoritis (distribusi atau prediksi) sesuai dengan data observasi nyata. Uji ini membandingkan frekuensi aktual (observed) yang tercatat dalam sampel dengan frekuensi harapan (expected) yang dihasilkan oleh model [27], menggunakan rumus *Chi-square*:

$$X^2 = \sum \frac{(O_i - E_i)^2}{E_i} \quad (6)$$

Keterangan :  $O_i$ : Frekuensi observasi (observed frequency) ke-I,  $E_i$ : Frekuensi harapan (expected frequency) ke-I,  $\sum$  : Penjumlahan untuk semua kategori atau kelas data. Hasil uji menentukan apakah perbedaan antara observasi dan ekspektasi signifikan secara statistik, yang mengindikasikan kecocokan model (good fit) jika tidak signifikan, atau ketidakcocokan (poor fit) jika signifikan [27].

### 2.6.2 Chi-square Test of Independence

*Chi-square Test of Independence* ini Uji ini mengevaluasi apakah dua atau lebih variabel kategorikal dalam suatu populasi bersifat independen (tidak terkait), dengan menganalisis ketidaksetaraan antara distribusi variabel dalam sampel acak [28]. Dapat dilihat pada Tabel 1., untuk tabel kontingensi *test of independence*:

Tabel 1. Tabel Kontingensi

Sentimen	Akuisisi Ya	Akuisisi Tidak	Total baris (i)
Positif	Jumlah	Jumlah	Total jumlah
Negatif	Jumlah	Jumlah	Total jumlah
Netral	Jumlah	Jumlah	Total jumlah
Total kolom (j)	Total Ya	Total Tidak	Jumlah total

Untuk perhitungan frekuensi harapan untuk *Chi-square test of independence* bisa dilihat sebagai berikut:

$$E_{ij} = \frac{\text{total baris}_i \times \text{total kolom}_j}{\text{total keseluruhan}} \quad (7)$$

$E_{ij}$  = frekuensi harapan pada sel baris ke-i, kolom ke-j

$\text{total baris}_i$  = jumlah total pada baris ke-i

$\text{total kolom}_j$  = jumlah total pada kolom ke-j

$\text{total keseluruhan}$  = total semua data

Untuk rumus perhitungan *Chi-square Test Of Independence* bisa dilihat dalam rumus ini:

$$X^2 = \sum_{i=1}^3 \sum_{j=2}^2 \frac{O_{ij} - E_{ij}}{E_{ij}} \quad (8)$$

Hasil uji ditentukan jika sampel menunjukkan korelasi yang signifikan, disimpulkan bahwa variabel tersebut kemungkinan besar tidak independen dalam populasi asal [28].

### 2.6.3 Odds Ratios

*Odds Ratio* adalah ukuran statistik yang menunjukkan seberapa kuat hubungan dan perbandingan peluang antara dua kelompok. *Odds ratio* diperlukan untuk melengkapi hasil uji *Chi-square* yang sudah didapat, memberikan ukuran efek untuk ada/tidaknya hubungan antar kelompok dan dapat diinterpretasikan dalam bahasa umum. Untuk rumus pencarian *Odds Ratio* ini dapat dilihat sebagai berikut:

$$OR = \frac{a \times d}{b \times c} \quad (9)$$

$a$  = kejadian positif dikelompok 1

$b$  = kejadian negatif dikelompok 1

$c$  = kejadian positif dikelompok 2

$d$  = kejadian negatif dikelompok 2

## 3. HASIL DAN PEMBAHASAN

### 3.1 Analisis Kebutuhan Data

Analisis ini bertujuan menemukan poin penting dalam penelitian dan merencanakan pelaksanaannya. Penelitian menggunakan data komentar dari Instagram Insyira Pekanbaru periode Mei 2024 sampai Mei 2025. Data ini dianalisis untuk memahami persepsi pelanggan dan seberapa efektif strategi pemasaran yang dilakukan.

### 3.2 Pengumpulan Data

Data penelitian ini diambil dari komentar pada unggahan Instagram @insyiraoleholehpekanbaru. Pengambilan data dilakukan menggunakan program *Python* dan *Extension Instant Data Scrapper* untuk mengumpulkan semua komentar. Total data seluruh komentar yang terkumpul selama satu tahun terakhir dari periode 20 Mei 2024 hingga 24 Mei 2025 adalah sebesar 1363 komentar. Dapat dilihat pada Dengan pengumpulan data ini, penelitian dapat mengkaji pendapat dan respons pelanggan secara lengkap, sehingga menghasilkan analisis yang tepat tentang persepsi konsumen terhadap produk/layanan Insyira Pekanbaru.

### 3.3 Preprocessing Data

Data komentar yang telah diambil melalui proses *scraping* akan menjalani tahap *preprocessing*. Tujuan dari *preprocessing* ini adalah untuk membersihkan data mentah yang telah diperoleh. Langkah pertama dalam *preprocessing* adalah melakukan *cleaning* data. *Cleaning* data bertujuan untuk menghapus emotikon, hashtag, dan mention yang ada dalam

komentar. Pada tahap *cleaning* ini, penulis melakukan pembersihan secara manual terhadap data yang mengandung *mention* dan *emotikon*. Untuk tahap ini penulis menggunakan *Python* dan *Regex* untuk *cleaning* teks untuk pembersihan symbol, emoticon dan spasi atau angka yang berlebih pada data komentar instagram. Proses ini bisa di lihat pada gambar 2:

```
def clean_text(text):
    text = re.sub(pattern=r'@\w+|#\w+|http\S+|\b\d+\b|^a-zA-Z\s', repl='', text, flags=re.IGNORECASE)
    text = re.sub(pattern=r'(\.){1,2}', repl=r'\1', text)
    return text.lower().strip()
```

**Gambar 1.** Penggunaan *RegEx* di *Python*

Lalu melakukan pemanggilan data yang berisi dengan seluruh komentar instagram yang ingin dilakukan *Preprocessing*. berikut Gambar 2 untuk pemanggilan file *excel* yang berisi komentar instagram.

```
df = pd.read_excel(io='INC_INSI.xlsx', sheet_name='Sheet1', usecols=['Column1', 'Column2'])
df = df.rename(columns={'Column1': 'Akun', 'Column2': 'Komentar'}).dropna(subset=['Komentar'])
df['Komentar'] = df['Komentar'].astype(str)
```

**Gambar 2** Pemanggilan File Excel

Hasil data table setelah pemanggilan bisa dilihat dalam tabel sudah dibersihkan dengan *RegEx* dapat dilihat dalam tabel 2:

**Tabel 2.** Data Asli

No	Komentar
1	Pisang frozen by lima jari terthe best 😊
2	Enak,praktis dan terjangkau 📦
3	Recomended buat oleh oleh ke sawahlunto 😊
4	Pisang kipas frozen itu berapa harga , msg2 ukuran..boleh tau Kak.
5	Gak hanya lebaran yg macet tp sore selalu bikin macet
6	Sangat lezat n nikmat
7	Alhamdulillah kami sudah pernah makannya
1363	Masya Allah 😊

Dapat dilihat dari Tabel diatas masih belum dilakukan proses *cleaning*. Lalu dengan *RegEx* akan dilakukan proses *cleaning* dan melakukan *case folding* atau pengubahan kata menjadi huruf kecil. Bisa dilihat dalam tabel bahwa simbol dan *emoticon* telah dibersihkan dan menjadi huruf kecil pada **Tabel 1** ini.

**Tabel 1.** Case Folding dan *RegEx*

No	Komentar
1	pisang frozen by lima jari ter the best
2	enak praktis dan terjangkau
3	recomended buat oleh oleh ke sawah lunto
4	pisang kipas frozen itu berapa harga msg ukuran boleh tau kak
5	gak hanya lebaran yg macet tp sore selalu bikin macet
6	sangat lezat n nikmat
7	alhamdulillah kami sudah pernah makan nya
1363	masya allah

Setelah tahap ini, akan dilakukan proses *Tokenizing* dan *Stemming* yang akan dilakukan di *Python* dan menggunakan *Library NLTK (Natural Language Toolkit)* dan juga menggunakan *stopword*.

### 3.3.1 Tokenization dan Stemming

*Tokenizing* berguna untuk memecah kalimat menjadi kata per kata. Pada tahap ini, libary *tokenizing* diambil langsung dari *NLTK (Natural Language Toolkit)*. Selanjutnya dilakukan proses *stopword* yang berfungsi untuk menghapus kata-kata yang bukan merupakan kata kunci pada data komentar. Pada tahap ini, libary *stopword* diambil dari *NLTK* dan tambahan repositori dari *github* yang dikembangkan menjadi *stopword-id* yang berupa kumpulan kata kunci yang ada di dalam bahasa indonesia. Lalu dilakukan *stemming* data, *stemming* berfungsi untuk menyelaraskan kalimat untuk



mengurangi daftar kata, menghilangkan imbuhan, menjadikan kata dasar. stemming ini menggunakan *Sastrawi* pada *python*. Semua proses ini dijalankan secara bersamaan dan dapat dilihat pada Gambar 3 berikut:

```

1 from nltk.corpus import stopwords
2 from nltk.tokenize import word_tokenize
3 from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

```

Gambar 3. Preprocess Data

Lalu dilakukan pemanggilan tabel yang sama dengan proses *python* diatas tadi dan akan menghasilkan Tabel 2 seperti berikut ini.

Tabel 2. Hasil Preprocess Data

No	Komentar	Tokenized stopword	Stemming
1	pisang frozen by lima jari ter the best	Pisang, frozen, by, lima, jari, the, best	Pisang, frozen, by, lima, jari, the, best
2	enak praktis dan terjangkau	Enak, praktis, dan, terjangkau	Enak, praktis, terjangkau
3	recomended buat oleh oleh ke sawah lunto	Recomended, buat, oleh, ke, sawahlunto	buat, oleh, ke, sawahlunto
4	pisang kipas frozen itu berapa harga msg ukuran boleh tau kak	Pisang, kipas, frozen, itu, berapa, harga, msg, ukuran, boleh, tau, kak	Pisang, kipas, frozen, berapa, harga, ukuran, boleh, kak
5	gak hanya lebaran yg macet tp sore selalu bikin macet	Tidak, hanya, lebaran, yang, macet, tapi, sore, selalu, bikin, macet	Tidak, lebaran, macet, sore, selalu, bikin, macet
6	sangat lezat n nikmat	Sangat, lezat, dan, nikmat	Lezat, nikmat
7	alhamdulillah kami sudah pernah makan nya	Alhamdulillah, kami, sudah, pernah, makan	Alhamdulillah, pernah, makan
1363	masya allah	Masyaallah	Masyaallah

### 3.4 Pelabelan

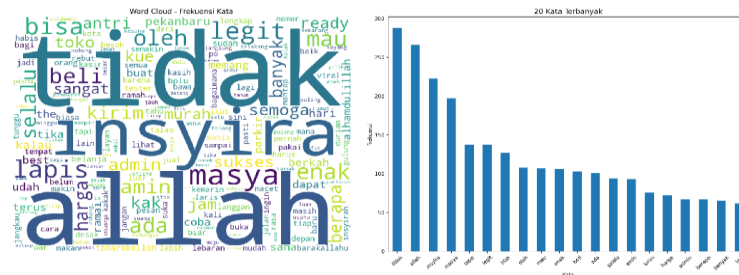
Pelabelan sentimen ada beberapa tahapan yang dimulai dengan identifikasi sentimen menggunakan lexicon, yang mendeteksi kata positif atau negatif. Selanjutnya, dilakukan analisis konteks menggunakan pendekatan *rule-based* dan *RegEx* untuk mengkaji adanya negasi atau ironi yang bisa memengaruhi makna. Tahap kurasi manual dilakukan untuk memastikan ketepatan label, terutama pada komentar yang ambigu. Lexicon digunakan sebagai dasar linguistik, sementara kurasi manual berperan dalam menjaga akurasi hasil akhir. Melalui tahapan ini, setiap komentar dapat diklasifikasikan secara tepat ke dalam kategori positif, negatif, atau netral, dan hasil ini dapat dilihat pada Tabel 3:

Tabel 3. Pelabelan Data

No	Komentar	Label sentimen
1	pisang frozen by lima jari the best	Positif
2	enak praktis dan terjangkau	Netral
3	recomended buat oleh oleh ke sawah lunto	Positif
4	pisang kipas frozen itu berapa harga msg ukuran boleh tau kak	Netral
5	gak hanya lebaran yg macet tp sore selalu bikin macet	Negatif
6	lezat nikmat	Positif
7	alhamdulillah pernah makan	Netral
1363	masya allah	Netral

### 3.5 Visualisasi Data

Visualisasi data adalah proses menyajikan informasi dalam bentuk grafis seperti grafik, diagram, atau gambar. Ini ditujukan agar pola, tren, dan hubungan dalam data lebih mudah dipahami. Tahapan penelitian ini menggunakan Word Cloud dan diagram batang Top 20 Kata untuk menampilkan kata-kata dominan berdasarkan frekuensi kemunculannya. Keduanya berfungsi mengidentifikasi kata kunci sentral dalam data teks, mempermudah pemahaman pola komunikasi pengguna, serta mengungkap fokus pembicaraan dan sentimen dominan secara sekilas. Visualisasi data dapat dilihat dari Gambar 4 sebagai berikut.



Gambar 4. Word Cloud

### 3.6 Uji Chi-square Goodness of Fit

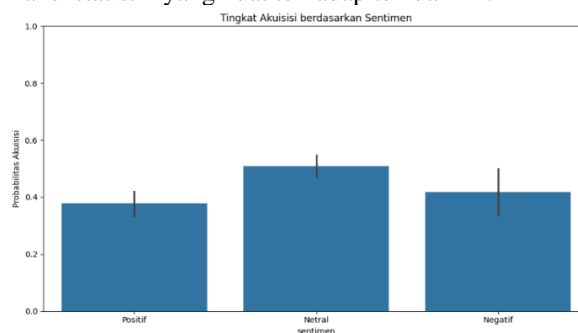
Hasil uji *Chi-square Goodness of Fit* menunjukkan perbedaan sangat signifikan antara distribusi sentimen aktual dan distribusi merata hipotetis ( $\chi^2 = 321.51$ ,  $p < 0.0001$ ). Dari total 1.363 komentar, teramati pola tidak merata: sentimen netral mendominasi (650 komentar atau 47.7%), diikuti positif (567 komentar atau 41.6%), dan negatif (146 komentar atau 10.7%). Nilai *Chi-square* tinggi (321.51) mengindikasikan perbedaan substansial antara frekuensi aktual dan frekuensi harapan (454.33 per kategori jika merata). P-value sangat kecil ( $p < 0.0001$ ) menegaskan bahwa pola ini bukan kebetulan, melainkan mencerminkan kecenderungan nyata dalam data. Ukuran efek *Cramer's V* (0.343) mengkonfirmasi ketidakmerataan signifikan, meski kekuatannya moderat. Dominasi sentimen netral (47.7%) dan positif (41.6%) menunjukkan kecenderungan pengguna memberikan komentar bersifat informatif dan apresiatif. Temuan ini konsisten dengan penelitian sebelumnya: komentar netral umumnya berisi pertanyaan teknis produk, sementara positif berupa apresiasi terhadap layanan.

### 3.7 Uji Chi-square Test of Independence

Hasil uji *Chi-square Test of Independence* menghasilkan hubungan sangat signifikan antara sentimen komentar dan perilaku akuisisi ( $\chi^2 = 21.78$ ,  $p < 0.0001$ ). Analisis tabel kontingensi terhadap 1.363 komentar menunjukkan pola: komentar netral mencapai tingkat akuisisi tertinggi (50.9% atau 331 dari 650 komentar), melebihi komentar positif (37.7% atau 214 dari 567 komentar) sebesar 13.2%, dan mengungguli komentar negatif (41.8% atau 61 dari 146 komentar). Dengan hasil ini, membuktikan bahwa komentar dengan sentiment netral lebih dekat dalam tolak ukur akuisisi sebagai sentimen paling penting untuk menghasilkan akuisisi atau membuat calon pelanggan tertarik untuk melakukan transaksi.

### 3.8 Uji Odds Ratio

Hasil uji *Odds Ratio* mengungkap hubungan *paradoksal* antara sentimen komentar dan perilaku akuisisi yang membalik asumsi normal. Analisis terhadap 1.363 komentar membuktikan bahwa komentar netral justru memiliki efektivitas akuisisi tertinggi, dengan komentar positif menunjukkan penurunan peluang akuisisi sebesar 42% dibandingkan netral ( $OR = 0.58$ ; 95%  $CI [0.46, 0.73]$ ;  $p < 0.0001$ ). *Confidence Interval* yang sempit dan eksklusif (tidak mencakup nilai netral 1) mengindikasikan signifikansi statistik yang kuat terhadap temuan ini.



Gambar 5. Uji Odds Ratio



Perhitungan matematis mendasar memperkuat konsistensi hasil: penerapan rumus  $(a \times d) / (b \times c)$  dengan nilai \*a\* (positif-akuisisi) = 214, \*b\* (positif-non akuisisi) = 353, \*c\* (netral-akuisisi) = 331, dan \*d\* (netral-non akuisisi) = 319 menghasilkan  $OR = (214 \times 319) / (353 \times 331) = 68.266 / 116.843 = 0.58$ . Hasil ini selaras dengan temuan uji *Chi-square* sebelumnya ( $\chi^2 = 21.78$ ; \*p\* < 0.0001) dan didukung ukuran efek Cramer's V sebesar 0.13 yang mengonfirmasi kekuatan hubungan signifikan meskipun dalam kategori rendah-sedang.

### 3.9 Hasil Analisis Statistik Sentimen

Berdasarkan analisis statistik terhadap 1.363 komentar pelanggan, distribusi sentimen menunjukkan mayoritas komentar bersifat netral (650 komentar atau 47.7%), diikuti positif (567 komentar atau 41.6%), dan negatif (146 komentar atau 10.7%). Persentase ini dihitung menggunakan rumus  $\text{Persentase Sentimen} = (\text{Jumlah Komentar per Kategori} / \text{Total Komentar}) \times 100$ . Lebih lanjut, tingkat akuisisi per kategori sentimen dihitung dengan rumus  $\text{Tingkat Akuisisi} = (\text{Jumlah Akuisisi per Sentimen} / \text{Total Komentar per Sentimen}) \times 100$ , menghasilkan: komentar netral mencapai 50.9% akuisisi (331 dari 650 komentar), positif 37.7% (214 dari 567), dan negatif 41.8% (61 dari 146). Secara keseluruhan, total akuisisi mencapai 44.5% (606 dari 1.363 komentar). Proporsi sumber akuisisi menunjukkan dominasi komentar netral (54.6% dari total akuisisi), disusul positif (35.3%), dan negatif (10.1%). Hal ini membuktikan bahwa meskipun komentar positif memiliki intensitas emosional lebih tinggi, komentar netral justru menjadi penyumbang akuisisi pelanggan terbesar dalam penelitian ini.

**Tabel 4.** Hasil Analisis Sentimen

Sentimen	Efektivitas akuisisi	Karakteristik Kunci
Netral	50.9%	Pertanyaan Produk, minat beli
Negatif	41.8%	Keluhan spesifik
Positif	37.7%	Interaksi tanpa target akuisisi

## 4. KESIMPULAN

Berdasarkan analisis 1.363 komentar Instagram @insyiraoleholehepekanbaru (Mei 2024-2025), distribusi sentimen didominasi komentar netral (47.7%), diikuti positif (41.6%) dan negatif (10.7%). Bukti statistik mengkonfirmasi hubungan kausal antara sentimen komentar Instagram dan akuisisi pelanggan pada Insyira Oleh-Oleh Pekanbaru. Hasil analisis menunjukkan bahwa komentar netral justru paling efektif mendorong akuisisi pelanggan (50.9%), melampaui komentar negatif (41.8%) dan positif (37.7%). Uji *Chi-square* menunjukkan bahwa perbedaan ini signifikan ( $\chi^2=21.78$ ;  $p<0.0001$ ). Hal ini dapat dijelaskan karena komentar netral sering berupa pertanyaan tentang produk yang menunjukkan minat beli aktif berbeda dari komentar positif yang cenderung bersifat pujian pasif tanpa niat transaksi. Temuan menunjukkan efektivitas akuisisi tertinggi pada komentar netral dengan mencapai tingkat akuisisi tertinggi, dengan pola setiap dua komentar netral menghasilkan satu akuisisi, didominasi pertanyaan tentang produk seperti yang menjadi faktor utama untuk pembelian. Sementara itu, komentar negatif menunjukkan efektivitas lebih tinggi (41.8%) dibanding positif (37.7%), didorong respons cepat tim CRM terhadap keluhan seperti "Barang habis" yang kerap berujung solusi dan konversi. Sebaliknya, komentar apresiatif ("Enak banget!") justru mencatat akuisisi terendah, mengindikasikan bahwa kepuasan pelanggan tidak secara otomatis berkonversi menjadi akuisisi pelanggan baru. Tiga pengujian yang sudah dilakukan menunjukkan hasil yang saling menguatkan dimana uji Goodness of Fit ( $\chi^2=321.51$ ; \*p\* < 0.0001) yang mengonfirmasi distribusi non-acak, Tabel Kontingensi dengan pola konsisten netral > negatif > positif, dan Odds Ratio (OR=0.58; CI[0.46,0.73]) yang mengindikasikan peluang akuisisi positif 42% lebih rendah dibanding komentar netral dan membentuk konsistensi triangular yang mengeliminasi keraguan. Temuan ini berbeda dengan asumsi umum dalam CRM : komentar netral, terutama yang berisi pertanyaan, justru memiliki potensi akuisisi lebih tinggi dibanding komentar positif. UMKM seperti Insyira oleh-oleh Pekanbaru perlu mengalihkan fokus strategi CRM ke arah respons cepat terhadap komentar jenis netral ini yang mayoritas itu pertanyaan.

## REFERENCES

- [1] S. Wan, C. Paris, And D. Georgakopoulos, "Social Media Data Aggregation And Mining For Internet-Scale Customer Relationship Management," In *Proceedings - 2015 Ieee 16th International Conference On Information Reuse And Integration, Iri 2015*, 2015, Pp. 39–48. Doi: 10.1109/Iri.2015.17.
- [2] S. Rawat, S. Jindal, R. S. Moorti, Y. Mangal, And N. Saxena, "Change In It World With The Evolution Of Social Media Using Big Data," In *2018 International Conference On Communication, Computing And Internet Of Things (Ic3iot)*, Chennai, India, 2018, Pp. 408–412. Doi: 10.1109/Ic3iot.2018.8668165.
- [3] Datareportal And S. Kemp, "Digital 2025: Indonesia," Datareportal – Global Digital Insights. Accessed: May 26, 2025. [Online]. Available: <https://Datareportal.Com/Reports/Digital-2025-Indonesia>
- [4] E. Retnoningsih, F. N. Khasanah, R. T. Handayanto, H. Herlawati, S. Rofiah, And S. Solikin, "Knowledge Management System For Supporting The Small Medium Enterprise (Ukm) In Bekasi City," In *Proceedings Of 2019 4th International Conference On Informatics And Computing, Icic 2019*, 2019, Pp. 1–5. Doi: 10.1109/Icic47613.2019.8985929.
- [5] R. H. M. Arbi Siti, A. Bismo, And L. Sutiyo, "Segmentation Analysis Of Instagram Users Based On Preferences Towards Forms



- And Types Of Online Marketing Content,” In *Proceedings Of 2019 International Conference On Information Management And Technology, Icimtech 2019*, Ieee, 2019, Pp. 202–207. Doi: 10.1109/Icimtech.2019.8843797.
- [6] R. Estela, J. Vasco, J. Carlos, P. Yuquilema, J. Fernando, And L. Aguirre, “Digital Marketing And Customer Acquisition In An Agro-Industrial Company,” *Russ. Law J.*, Vol. 11, No. 8, Pp. 774–787, 2023, Doi: 10.52783/Rlj.V11i8s.1420.
- [7] A. Chorianopoulos, “Classification Modeling Methodology,” In *Effective Crm Using Predictive Analytics*, John Wiley & Sons, 2015, Pp. 21–111. Doi: 10.1002/9781119011583.Ch2.
- [8] M. Tavakoli, M. Molavi, V. Masoumi, M. Mobini, S. Etemad, And R. Rahmani, “Customer Segmentation And Strategy Development Based On User Behavior Analysis, Rfm Model And Data Mining Techniques: A Case Study,” In *Proceedings - 2018 Ieee 15th International Conference On E-Business Engineering, Icebe 2018*, Ieee, 2018, Pp. 119–126. Doi: 10.1109/Icebe.2018.00027.
- [9] D. D. A. Yani, H. S. Pratiwi, And H. Muhardi, “Implementasi Web Scraping Untuk Pengambilan Data Pada Situs Marketplace,” *J. Sist. Dan Teknol. Inf.*, Vol. 7, No. 4, P. 257, 2019, Doi: 10.26418/Justin.V7i4.30930.
- [10] R. Vatrappu, R. R. Mukkamala, A. Hussain, And B. Flesch, “Social Set Analysis: A Set Theoretical Approach To Big Data Analytics,” *Ieee Access*, Vol. 4, Pp. 2542–2571, 2016, Doi: 10.1109/Access.2016.2559584.
- [11] Z. Madhoushi, A. R. Hamdan, And S. Zainudin, “Sentiment Analysis Techniques In Recent Works,” In *Proceedings Of The 2015 Science And Information Conference, Sai 2015*, 2015, Pp. 288–291. Doi: 10.1109/Sai.2015.7237157.
- [12] A. Kumar And S. P. Panda, “A Survey: How Python Pitches In It-World,” In *Proceedings Of The International Conference On Machine Learning, Big Data, Cloud And Parallel Computing: Trends, Prespectives And Prospects, Comitcon 2019*, Ieee, 2019, Pp. 248–251. Doi: 10.1109/Comitcon.2019.8862251.
- [13] D. Cielen, A. D. B. Meysman, And M. Ali, *Introducing Data Science: Big Data, Machine Learning, And More, Using Python Tools*. Manning Publications, 2016. [Online]. Available: <https://www.manning.com/books/introducing-data-science%0a>
- [14] M. Heidari And P. Shamsinejad, “Producing An Instagram Dataset For Persian Language Sentiment Analysis Using Crowdsourcing Method,” In *Proceedings - 2020 6th International Conference On Web Research (Icwr)*, 2020, Pp. 284–287. Doi: 10.1109/Icwr49608.2020.9122270.
- [15] N. I. Darmawan, M. K. Pratiwi, S.T., And B. D. Setiawan, “Analisis Sentimen Terhadap Akuisisi Saham Tokopedia Oleh Tiktok Menggunakan Naive Bayes Berdasarkan Komentar Youtube,” Universitas Brawijaya, 2017. [Online]. Available: <https://repository.ub.ac.id/Id/Eprint/227868/%0a>
- [16] Q. F. F. Safiesza, M. Afdal, And R. Novita, “Sentimen Analisis Social Crm Pada Media Sosial Instagram Menggunakan Machine Learning Untuk Mengukur Retensi Pelanggan,” Vol. 6, No. 2, Pp. 606–616, 2024, Doi: 10.47065/Bits.V6i2.5269.
- [17] B. Classifier, A. Erfina, M. Dholah, R. Ndk, And H. Ramdhan, “Indonesian Twitter Sentiment Analysis Application On The Covid19 Vaccine Using Php And Naive,” *Int. Conf. Comput. Eng. Des.*, Pp. 1–6, 2021, Doi: 10.1109/Icced53389.2021.9664864.
- [18] F. Levina Marthina Rumaropen, Mahananto, “Evaluation To Implementation Of Customer Relationship Management In Scope Of Social Media To Increase Interest Customer,” *J. Ilm. Manaj. Ubhara*, Vol. 7, No. 1, 2025, Doi: <https://doi.org/10.31599/jmu.v5i1>.
- [19] A. Cahya Kamilla, N. Priyani, R. Priskila, And V. Handrianus Pranatawijaya, “Analisis Sentimen Film Agak Laen Dengan Kecerdasan Buatan: Text Mining Metode Naive Bayes Classifier,” *Jati (Jurnal Mhs. Tek. Inform.)*, Vol. 8, No. 3, Pp. 2923–2928, 2024, Doi: 10.36040/Jati.V8i3.9587.
- [20] S. R. S. Kusumo, “Penerapan Web Scraping Deskripsi Produk Menggunakan Selenium Python Dan Framework Laravel,” *Jatissi (Jurnal Tek. Inform. Dan Sist. Informasi)*, Vol. 9, No. 4, Pp. 3426–3435, 2022, Doi: 10.35957/Jatissi.V9i4.2727.
- [21] R. A. Firsttama, A. A. Arifiyanti, And D. S. Y. Kartika, “Analisis Sentimen Komentar Youtube Konferensi Tingkat Tinggi G20 Menggunakan Metode Naive Bayes,” *J. Teknol. Dan Sist. Inf. Bisnis*, Vol. 6, No. 2, Pp. 282–285, 2024, Doi: 10.47233/Jteksis.V6i2.1263.
- [22] E. E. Amelia And I. Yustiana, “Analisis Sentimen Pada Ulasan Produk Uniqlo Dengan Algoritma Naive Bayes,” *J. Sains Komput. Inform.*, Vol. 8, No. 1, Pp. 141–148, 2024, Doi: <http://dx.doi.org/10.30645/J-Sakti.V8i1.773>.
- [23] 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.
- [24] F. T. Saputra, S. H. Wijaya, Y. Nurhadryani, And D. Defina, “Lexicon Addition Effect On Lexicon-Based Of Indonesian Sentiment Analysis On Twitter,” In *Proceedings - 2nd International Conference On Informatics, Multimedia, Cyber, And Information System, Icimcis 2020*, 2020, Pp. 136–141. Doi: 10.1109/Icimcis51567.2020.9354269.
- [25] F. Koto And G. Y. Rahmanyngtyas, “Inset Lexicon: Evaluation Of A Word List For Indonesian Sentiment Analysis In Microblogs,” In *Proceedings Of The 2017 International Conference On Asian Language Processing, Ialp 2017*, 2017, Pp. 391–394. Doi: 10.1109/Ialp.2017.8300625.
- [26] D. Jurafsky And J. H. Martin, *Speech And Language Processing: An Introduction To Natural Language Processing, Computational Linguistics, And Speech Recognition With Language Models*, 3rd Editio. Stanford University (Online Draft), 2025. Doi: 10.9783/9780812200027.Toc.
- [27] Y. Qiu, L. Liu, X. Lai, And Y. Qiu, “An Online Test For Goodness-Of-Fit In Logistic Regression Model,” *Ieee Access*, Vol. 7, Pp. 107179–107187, 2019, Doi: 10.1109/Access.2019.2927035.
- [28] N. S. Turhan, “Karl Pearsons Chi-Square Tests,” *Educ. Res. Rev.*, Vol. 15, No. 9, Pp. 575–580, 2020, Doi: 10.5897/Err2019.3817.