

***OPINION MINING QRIS PADA TWITTER
MENGUNAKAN VADER DAN LDA***

SKRIPSI



**UIN SUNAN AMPEL
S U R A B A Y A**

Disusun Oleh:

**Dzakiya Ishmatul Ulya
H96219045**

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SUNAN AMPEL
SURABAYA
2023**

PERNYATAAN KEASLIAN

Saya yang bertanda tangan di bawah ini,

Nama : DZAKIYA ISHMATUL ULYA
NIM : H96219045
Program Studi : Sistem Informasi
Angkatan : 2019

Menyatakan bahwa saya tidak melakukan plagiat dalam penulisan skripsi saya yang berjudul: “*OPINION MINING QRIS PADA TWITTER MENGGUNAKAN VADER DAN LDA*”. Apabila suatu saat nanti terbukti saya melakukan tindakan plagiat, maka saya bersedia menerima sanksi yang telah ditetapkan.

Demikian pernyataan keaslian ini saya buat dengan sebenar-benarnya.

Surabaya, 10 April 2023
Yang menyatakan,



The image shows a handwritten signature in black ink over a red and white revenue stamp. The stamp is a 1000 Rupiah stamp with the Garuda Pancasila emblem and the text 'SPUICR HBBU KUPAH 1000 Rp. METERAI TEMPEL' and the serial number '692C1AKX407848794'.

Dzakiya Ishmatul Ulya
NIM. H96219045

LEMBAR PERSETUJUAN PEMBIMBING


Skripsi oleh

NAMA : DZAKIYA ISHMATUL ULYA
NIM : H96219045
JUDUL : *OPINION MINING* QRIS PADA TWITTER
MENGUNAKAN VADER DAN LDA

ini telah diperiksa dan disetujui untuk diujikan.


Surabaya, 10 April 2023

Dosen Pembimbing I



Dr. Eng. Anang Kunaefi, M. Kom
NIP. 197911132014031001

Dosen Pembimbing II



Dwi Rolliawati, MT
NIP. 197909272014032001

PENGESAHAN TIM PENGUJI SKRIPSI

Skripsi Dzakiya Ishmatul Ulya ini telah dipertahankan
di depan tim penguji skripsi
di Surabaya, 17 April 2023

Mengesahkan,
Dewan Penguji

Penguji I



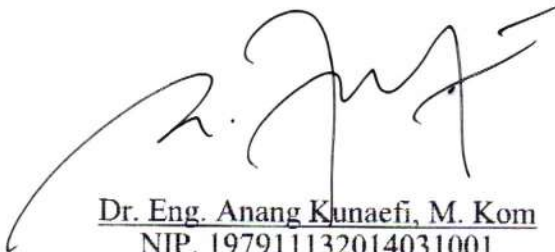
Bayu Adhi Nugroho, Ph.D.
NIP. 197905182014031001

Penguji II



Khalid, M. Kom
NIP. 197906092014031002

Penguji III



Dr. Eng. Anang Kunaefi, M. Kom
NIP. 197911132014031001


Penguji IV



Dwi Rolliawati, MT
NIP. 197909272014032001

Mengetahui,
Dekan Fakultas Sains dan Teknologi
Sunan Ampel Surabaya




Saepul Hamdani, M.Pd.
196507312000031002



UIN SUNAN AMPEL
SURABAYA

KEMENTERIAN AGAMA
UNIVERSITAS ISLAM NEGERI SUNAN AMPEL SURABAYA
PERPUSTAKAAN

Jl. Jend. A. Yani 117 Surabaya 60237 Telp. 031-8431972 Fax.031-8413300
E-Mail: perpus@uinsby.ac.id

LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI
KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademika UIN Sunan Ampel Surabaya, yang bertanda tangan di bawah ini, saya:

Nama : Dzakiya Ishmatul Ulya
NIM : H96219045
Fakultas/Jurusan : Sains dan Teknologi/Sistem Informasi
E-mail address : dishmatul@gmail.com

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Perpustakaan UIN Sunan Ampel Surabaya, Hak Bebas Royalti Non-Eksklusif atas karya ilmiah :

Sekripsi Tesis Desertasi Lain-lain (.....)

yang berjudul :

OPINION MINING QRIS PADA TWITTER MENGGUNAKAN VADER DAN LDA

beserta perangkat yang diperlukan (bila ada). Dengan Hak Bebas Royalti Non-Eksklusif ini Perpustakaan UIN Sunan Ampel Surabaya berhak menyimpan, mengalih-media/format-kan, mengelolanya dalam bentuk pangkalan data (database), mendistribusikannya, dan menampilkan/mempublikasikannya di Internet atau media lain secara *fulltext* untuk kepentingan akademis tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan atau penerbit yang bersangkutan.

Saya bersedia untuk menanggung secara pribadi, tanpa melibatkan pihak Perpustakaan UIN Sunan Ampel Surabaya, segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak Cipta dalam karya ilmiah saya ini.

Demikian pernyataan ini yang saya buat dengan sebenarnya.

Surabaya, 28 April 2023

Penulis

(Dzakiya Ishmatul Ulya)

ABSTRAK

OPINION MINING QRIS PADA TWITTER MENGUNAKAN VADER DAN LDA

Oleh:

Dzakiya Ishmatul Ulya

QRIS merupakan kode QR terstandarisasi yang dikembangkan oleh Bank Indonesia untuk memudahkan transaksi *mobile payment* di Indonesia. Adopsi penggunaan QRIS yang semakin meningkat menyebabkan banyaknya opini publik dan pengalaman masyarakat yang dibagikan di media sosial, termasuk Twitter. Penelitian ini bertujuan untuk mengetahui sentimen masyarakat dan topik dominan terkait QRIS di Indonesia menggunakan metode VADER dan LDA. Hasil penelitian mengungkap bahwa sebanyak 67% sentimen masyarakat terhadap QRIS adalah positif, 28,2% sentimen negatif, dan 4,17% sentimen netral dengan akurasi pelabelan sentimen sebesar 81,66%. Selain itu, terdapat enam topik dominan yang muncul pada *tweets* positif dengan C_V 0.488037, serta tiga topik dominan pada *tweets* negatif dengan C_V 0.383938. Hal ini membuktikan bahwa penggunaan QRIS di Indonesia masih mendapatkan respon positif dari masyarakat. Meskipun hasil analisis menunjukkan lebih banyak topik dengan sentimen positif daripada negatif, topik-topik negatif yang teridentifikasi dapat memberikan wawasan berharga guna meningkatkan sistem QRIS. Oleh karena itu, sangat penting bagi para *stakeholders* untuk memperhatikan *feedback* positif dan negatif untuk memastikan pengembangan dan keberhasilan berkelanjutan QRIS di Indonesia.

Kata Kunci: *Natural Language Processing, Analisis Sentimen, Opinion Mining, Pemodelan Topik, Mobile Payment, QRIS*

ABSTRACT

OPINION MINING QRIS ON TWITTER USING VADER AND LDA

By:
Dzakiya Ishmatul Ulya

QRIS is a standardized QR code developed by Bank Indonesia to facilitate mobile payment transactions in Indonesia. The increasing adoption of QRIS has led to many public opinions and experiences shared on social media, including Twitter. This study uses the VADER and LDA methods to determine the sentiment of the Indonesian public and the dominant topics related to QRIS. The study's results revealed that 67% of public sentiment towards QRIS was positive, 28.2% negative, and 4.17% neutral, with a sentiment labeling accuracy of 81.66%. Additionally, six dominant topics appear in positive tweets with a C_V of 0.488037, and three dominant topics in negative tweets with a C_V of 0.383938. This proves that the use of QRIS in Indonesia still receives a positive response from the public. Although the analysis results show more topics with positive sentiment than negative, the identified negative topics can provide valuable insights to improve the QRIS system. Therefore, it is crucial for stakeholders to pay attention to both positive and negative feedback to ensure the development and sustainable success of QRIS in Indonesia.

Keywords: *Natural Language Processing, Sentiment Analysis, Opinion Mining, Topic Modeling, Mobile Payment, QRIS*

DAFTAR ISI

Halaman Judul	i
PERNYATAAN KEASLIAN	i
LEMBAR PERSETUJUAN PEMBIMBING	ii
PENGESAHAN TIM PENGUJI SKRIPSI	iii
PERNYATAAN PERSETUJUAN PUBLIKASI	iv
UCAPAN TERIMA KASIH	v
KATA PENGANTAR	vi
ABSTRAK	vii
ABSTRACT	viii
DAFTAR ISI	ix
DAFTAR GAMBAR	xii
DAFTAR TABEL	xiii
BAB I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Perumusan Masalah.....	4
1.3. Batasan Masalah.....	4
1.4. Tujuan Penelitian.....	4
1.5. Manfaat Penelitian.....	4
BAB II TINJAUAN PUSTAKA	6
2.1. Tinjauan Penelitian Terdahulu	6
2.2. Teori-teori Dasar	9
2.2.1. QRIS.....	9
2.2.2. <i>Text Mining</i>	9

2.2.3.	Analisis Sentimen	14
2.2.4.	Pemodelan Topik	17
2.2.5.	<i>Library</i> yang digunakan	21
2.3.	Integrasi Keilmuan	22
BAB III METODE PENELITIAN		25
3.1.	Tahapan Penelitian	25
3.1.1.	Perumusan Masalah	25
3.1.2.	Studi Literatur	26
3.1.3.	<i>Data Collection</i>	26
3.1.4.	<i>Data Cleaning</i>	27
3.1.5.	<i>Data Preprocessing</i>	27
3.1.6.	Analisis Sentimen	31
3.1.7.	Pemodelan Topik	32
BAB IV HASIL DAN PEMBAHASAN		35
4.1.	<i>Data Collection</i>	35
4.2.	<i>Data Cleaning</i>	35
4.3.	<i>Data Preprocessing</i>	36
4.3.1.	<i>Remove Username, Hashtag, URLs</i>	36
4.3.2.	<i>Remove Duplicates</i>	36
4.3.3.	<i>Remove Punctuation</i>	36
4.3.4.	<i>Remove Emoji</i>	37
4.3.5.	<i>Case Folding</i>	37
4.3.6.	<i>Normalization</i>	38
4.3.7.	<i>Tokenization</i>	38
4.3.8.	<i>Remove Stopwords</i>	39
4.3.9.	<i>Drop Tweets</i>	40

4.3.10. <i>Stemming</i>	40
4.4. Analisis Sentimen.....	41
4.4.1. Pelabelan Sentimen	41
4.4.2. <i>Data Validation</i>	41
4.4.3. <i>Evaluation</i>	43
4.4.4. Analisis Hasil Sentimen	44
4.5. Pemodelan Topik.....	49
BAB V PENUTUP	57
5.1. Kesimpulan.....	57
5.2. Saran.....	57
DAFTAR PUSTAKA	59



UIN SUNAN AMPEL
S U R A B A Y A

DAFTAR GAMBAR

Gambar 2.1 Notasi <i>Plate</i> LDA (Blei et al., 2003).....	19
Gambar 3.1 Diagram Alir Langkah-langkah Penelitian	25
Gambar 3.2 Diagram Alir <i>Data Cleaning</i>	27
Gambar 3.3 Diagram Alir <i>Data Preprocessing</i>	27
Gambar 3.4 Performa <i>Dictionary Matching</i> (Jha and Mahmoud, 2019).....	30
Gambar 3.5 Diagram Alir Klasifikasi Sentimen.....	31
Gambar 3.6 Diagram Alir <i>Data Validation & Evaluation</i>	31
Gambar 3.7 Diagram Alir Pemodelan Topik.....	32
Gambar 4.1 <i>Confusion Matrix</i> Hasil Evaluasi	43
Gambar 4.2 Distribusi Sentimen QRIS.....	44
Gambar 4.3 Tren QRIS Selama Tiga Tahun.....	45
Gambar 4.4 <i>Wordcloud</i> Sentimen Positif	46
Gambar 4.5 <i>Wordcloud</i> Sentimen Negatif.....	48
Gambar 4.6 Hasil Pemodelan Topik <i>Tweets</i> Positif.....	50
Gambar 4.7 Hasil Pemodelan Topik <i>Tweets</i> Negatif.....	53

UIN SUNAN AMPEL
S U R A B A Y A

DAFTAR TABEL

Tabel 2.1 Tinjauan Penelitian Terdahulu	6
Tabel 2.2 <i>Confusion Matrix</i>	13
Tabel 2.3 Perbandingan <i>F1 Score</i> VADER (Hutto and Gilbert, 2014).....	15
Tabel 2.4 Cuplikan Daftar Kata pada InSet <i>Lexicon</i>	17
Tabel 3.1 Dataset <i>Tweets</i> dengan <i>Keyword QRIS</i>	26
Tabel 3.2 Ilustrasi <i>Tweets</i> Duplikat dengan <i>Link URL</i>	28
Tabel 3.3 Colloquial <i>Lexicon</i>	29
Tabel 3.4 <i>Stemming</i> dengan Sastrawi (“Sastrawi,” 2023).....	30
Tabel 3.5 Proses Konversi <i>Dictionary</i>	32
Tabel 3.6 <i>Dictionary</i> ke <i>Bag of Words</i>	32
Tabel 4.1 Jumlah <i>Tweets</i>	35
Tabel 4.2 Hasil <i>Remove Username, Hashtag, URLs</i>	36
Tabel 4.3 Hasil <i>Remove Punctuation</i>	36
Tabel 4.4 Hasil <i>Remove Emoji</i>	37
Tabel 4.5 Hasil <i>Case Folding</i>	38
Tabel 4.6 Hasil <i>Normalization</i>	38
Tabel 4.7 Hasil <i>Tokenization</i>	39
Tabel 4.8 Hasil <i>Remove Stopwords</i>	39
Tabel 4.9 Hasil <i>Stemming</i>	40
Tabel 4.10 Hasil Pelabelan.....	41
Tabel 4.11 <i>Data Validation</i>	42
Tabel 4.12 Hasil Evaluasi	44
Tabel 4.13 Daftar 10 Kata Teratas Sentimen Positif.....	47
Tabel 4.14 Daftar 10 Kata Teratas Sentimen Negatif	48
Tabel 4.15 Tabel Kata “Tidak Bisa”	49
Tabel 4.16 Daftar Koherensi Pemodelan Topik <i>Tweets</i> Positif.....	50
Tabel 4.17 Daftar Koherensi Pemodelan Topik <i>Tweets</i> Negatif.....	53

BAB I

PENDAHULUAN

1.1. Latar Belakang

Teknologi *mobile payment* tersebar secara pesat ke seluruh dunia. Dompot digital dan mata uang digital diprediksi akan mengalihkan kartu debit dan kartu kredit (Pal et al., 2021). Alipay dan Wechat Pay telah menguasai lebih dari 90% *market share* di sektor *mobile payment* dengan lebih dari 500 juta pengguna. Meluasnya penggunaan *mobile payment* di China telah memberikan gambaran tentang bagaimana masa depan tanpa uang tunai (*cashless*) (Aveni and Roest, 2017). *Contactless payment* dan *mobile payment* menjadi alternatif bagi masyarakat yang mengkhawatirkan risiko kesehatan dari penggunaan uang kertas selama pandemi *COVID-19*. Lebih dari 27% UMKM di Amerika Serikat melaporkan peningkatan layanan pembayaran seluler selama pandemi (Balch, 2020).

Di Indonesia, terdapat sejumlah *mobile payment* berupa dompet digital yang banyak digunakan dan populer di berbagai kalangan masyarakat, seperti Gopay, Dana, Ovo, dan Shopeepay. Keunggulan dari penggunaan dompet digital ini terletak pada kemudahan, kenyamanan, dan keamanannya. Namun saat bertransaksi, penjual atau *merchant* harus menyediakan QR Code berasal dari masing-masing Penyelenggara Jasa Sistem Pembayaran (PJSP) yang sama. Munculnya berbagai PJSP baru membuat sistem QR Code menjadi *silo* dan tidak efektif bagi *merchant* dan pelanggan.

Dikarenakan ketidakefektifan dalam bertransaksi yang mengharuskan menggunakan QR Code masing-masing PSJP, Bank Indonesia sebagai otoritas yang mengatur Gerbang Pembayaran Nasional (GPN) merancang sebuah sistem yang akan mengintegrasikan berbagai jenis PJSP. Bank Indonesia telah menciptakan kode QR terstandarisasi guna memudahkan dan memfasilitasi transaksi *mobile payment* di Indonesia bernama QRIS (*Quick Response Code Indonesian Standard*). “Kode QR terstandarisasi ini dikembangkan bersama-sama dengan Asosiasi Sistem Pembayaran Indonesia (ASPI) memiliki tujuan agar dapat mendorong efisiensi transaksi, mempercepat inklusi keuangan, memajukan UMKM, dan harapannya dapat mendorong pertumbuhan ekonomi” ucap Kepala Perwakilan BI Provinsi DKI Jakarta, Onny Widjanarko saat peresmian QRIS.

QRIS merupakan satu kode QR terstandarisasi yang dapat digunakan segala jenis transaksi digital (Saputri, 2020). Dengan gagasan integrasi QR *Code* antar PSJP, QRIS dikembangkan guna memenuhi tahapan dalam implementasi visi Sistem Pembayaran Indonesia (SPI) pada tahun 2025. Dilansir dari qris.id, saat ini (14/01/2023) QRIS telah beroperasi di 416 Kabupaten dan 98 Kota di seluruh penjuru Indonesia, dipercaya oleh 129.045 *merchant* dari seluruh Nusantara, dan berhasil mengelola data transaksi QRIS sebesar Rp391.717.077.567 dari semua *merchant*. Gubernur Bank Indonesia Perry Warjiyo menyampaikan, Indonesia telah mengkonfirmasi komitmen empat negara ASEAN dalam mengoneksikan sistem pembayaran dengan QRIS, beberapa negara tetangga di antaranya yakni Singapura, Thailand, Filipina, dan Malaysia (TimTVOne, 2022).

Banyak pihak terbantu dengan adanya QRIS dalam memudahkan bertransaksi sehingga banyak juga masyarakat Indonesia yang berbagi pengalaman mereka saat menggunakan QRIS di media sosial. Media sosial seperti Twitter, Instagram, atau Facebook telah menjadi saluran komunikasi yang populer dalam dekade terakhir bagi banyak individu untuk mengekspresikan perasaan, pikiran, dan ide mereka. Tidak ada batasan yang ketat pada media sosial, memungkinkan siapapun untuk menuliskan hampir segala sesuatu mencakup semua aspek dan topik kehidupan mulai dari kehidupan sehari-hari, politik, kesehatan, pendidikan, hiburan, *consumer goods*, layanan dan masih banyak lagi. Pada tahun 2022, Indonesia termasuk lima negara dengan pengguna Twitter aktif terbanyak sebesar 4,23% yaitu 18,45 juta pengguna dari total pengguna Twitter sebanyak 436 juta di seluruh belahan dunia (Statista, 2022). Didukung dengan terjadi peningkatan jumlah pengguna sebesar 31,3% dibandingkan tahun sebelumnya (Monavia, 2022). Berdasarkan banyaknya jumlah pengguna Twitter aktif di Indonesia dataset dari *platform* Twitter dapat digunakan sebagai dataset dalam penelitian ini.

Dengan banyaknya pengguna Twitter di Indonesia, akan sangatlah menarik dan bermanfaat untuk menganalisis sentimen masyarakat terhadap topik tertentu. Analisis sentimen atau juga dikenal sebagai *opinion mining*, merupakan suatu bidang yang mengkaji analisis opini, perasaan, penilaian, emosi dan sikap seseorang terhadap suatu entitas seperti permasalahan, layanan, produk, individu, organisasi, acara, dan topik (Liu, 2012). Analisis sentimen secara umum dapat

dilakukan dengan menggunakan pendekatan *supervised machine learning* atau pendekatan *rule-based* berbasis *lexicon* (Albrecht et al., 2020). Terdapat beberapa metode analisis sentimen berbasis *lexicon*, salah satu metode yang sering digunakan adalah VADER (*Valence Aware Dictionary for sEntiment Reasoning*) (D, 2019; Ekaputri and Akbar, 2022; Mathayomchan et al., 2022; Rintyarna, 2021; Shurrab et al., 2021; Wedel et al., 2022; Yin et al., 2022). Dalam melakukan analisis sentimen dengan dataset berupa data *tweets*, VADER menghasilkan tingkat akurasi tertinggi dengan skor sebesar 0,96 ketika dibandingkan dengan 11 metode lainnya (Hutto and Gilbert, 2014).

Terdapat banyak sekali bidang kajian analisis sentimen, salah satunya adalah analisis sentimen berbasis topik. Analisis sentimen berbasis topik memberikan dasar untuk mengelola opini publik *online* dan memainkan peran penting dalam mengidentifikasi subjek populer dan perubahan sentimen di internet (Ali et al., 2022). *Topic Modeling* adalah metode *unsupervised learning* yang mengasumsikan bahwa setiap suatu dokumen merupakan kumpulan dari campuran topik dan pada setiap topiknya merupakan distribusi probabilitas atas kata-kata (Liu, 2012). *Latent Dirichlet Allocation* (LDA) dianggap sebagai metode yang fleksibel dan adaptif untuk *topic modeling* (Vayansky and Kumar, 2020). Sudah banyak penelitian terdahulu yang mengombinasikan metode *lexicon-based* seperti VADER dengan *topic modeling* LDA dalam menganalisis teks (Mathayomchan et al., 2022; Mohamed Ridhwan and Hargreaves, 2021; Shurrab et al., 2021; Wedel et al., 2022; Yiran and Srivastava, 2019). *Topic Modeling* menggunakan LDA dilakukan dalam penelitian ini agar dapat mengekstrak dan mengetahui topik apa saja yang dibicarakan masyarakat ketika beropini, menilai, dan berbagi pengalaman mereka saat menggunakan QRIS, berdasarkan *tweets* yang telah dilabeli menggunakan VADER.

Penelitian ini dapat digunakan untuk mengetahui opini masyarakat tentang QRIS, serta untuk mengevaluasi efektivitas implementasi standar ini. Dengan demikian, hasil dari penelitian ini harapannya dapat membantu dalam memberikan saran perbaikan kepada Bank Indonesia dalam meningkatkan sistem pembayaran dengan menggunakan QRIS.

1.2. Perumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, poin-poin permasalahan yang akan diselesaikan pada penelitian ini yakni:

1. Bagaimana hasil analisis sentimen QRIS menggunakan VADER?
2. Bagaimana hasil analisis pemodelan topik QRIS menggunakan *Latent Dirichlet Allocation* (LDA) berdasarkan pelabelan sentimen menggunakan VADER?

1.3. Batasan Masalah

Supaya dapat lebih berfokus, membahas secara rinci permasalahan dan tidak keluar pada cakupan, berikut batasan masalah penelitian:

1. Data merupakan *tweets* berbahasa Indonesia dengan kata kunci “QRIS” dan *tweets* yang tersedia sejak QRIS dirilis, 17 Agustus 2019 hingga 17 Agustus 2022.
2. Data *tweets* yang diolah hanya *tweets* yang memiliki ≥ 12 kata (Jha and Mahmoud, 2019).
3. Data *tweets* yang digunakan dalam analisis hanya *tweets* yang berlabel positif dan negatif.

1.4. Tujuan Penelitian

Dari narasi latar belakang dan permasalahan yang telah dirumuskan, tujuan penelitian ini sebagai berikut:

1. Mengetahui hasil analisis sentimen masyarakat di media sosial Twitter terkait QRIS menggunakan VADER.
2. Mengetahui hasil analisis topik dominan yang ditemukan dalam pemodelan topik di media sosial Twitter terkait QRIS menggunakan *Latent Dirichlet Allocation* (LDA) berdasarkan pelabelan sentimen menggunakan VADER.

1.5. Manfaat Penelitian

Agar dapat berkontribusi dan memberi dampak positif. Manfaat dari penelitian ini sebagai berikut:

1. Manfaat dari segi akademis
 - a. Memberikan kontribusi pada pengembangan metode analisis sentimen dan deteksi topik pada data *tweets* yang berkaitan dengan QRIS.

- b. Penelitian ini menjadi referensi bagi penelitian sejenis yang ingin mengkaji sentimen dan topik pada data *tweets* yang berkaitan dengan QRIS.
2. Manfaat dari segi praktis/aplikatif
- a. Pihak yang berkepentingan dapat memantau dan mengelola opini publik di media sosial Twitter terkait QRIS.
 - b. Penelitian ini dapat memberikan saran perbaikan bagi pihak yang berkepentingan dalam meningkatkan kualitas layanan QRIS.
 - c. Dapat digunakan dan menjadi rujukan sebagai dasar untuk mengambil keputusan (*data-driven decision making*) dalam pengembangan layanan QRIS agar dapat lebih baik dan tepat sesuai dengan kebutuhan masyarakat.



UIN SUNAN AMPEL
S U R A B A Y A

BAB II

TINJAUAN PUSTAKA

2.1. Tinjauan Penelitian Terdahulu

Melakukan peninjauan pada penelitian-penelitian terdahulu dapat menunjang dan menambah pemahaman topik penelitian. Dibuatlah Tabel 2.1 berupa tinjauan penelitian terdahulu beserta korelasi dengan penelitian yang akan dilakukan.

Tabel 2.1 Tinjauan Penelitian Terdahulu

No.	Judul	Hasil	Korelasi
1.	“ <i>Sentiment analysis and topic modeling for COVID-19 vaccine discussions</i> ” (Yin et al., 2022)	Sebanyak 75.665 <i>tweets</i> terkait vaksin <i>COVID-19</i> dianalisis untuk pelabelan sentimen menggunakan VADER, yang selanjutnya dimodelkan topik setiap label positif dan negatif. Penelitian ini mengupas sentimen masyarakat berdasarkan lokasi geografis, sentimen terhadap <i>brand-brand</i> vaksin, hingga umpan balik terkait vaksin.	Mengombinasikan metode VADER dan LDA dalam menganalisis <i>tweets</i> . VADER dipilih karena memiliki akurasi yang lebih tinggi dibanding metode lain.
2.	“ <i>Evolution of COVID-19 tweets about Southeast Asian Countries: topic modelling and sentiment analyses</i> ” (Mathayomchan et al., 2022)	Pemodelan topik LDA dilakukan dengan 115.553 <i>tweets</i> terkait pandemi <i>COVID-19</i> di 9 negara Asia Tenggara untuk mengidentifikasi 12 topik yang dikelompokkan menjadi tiga tema. Peneliti melakukan sentimen analisis dan deteksi emosi pada <i>tweets</i> berdasarkan topik yang telah didefinisikan.	
3.	“ <i>A Bilingual Comparison of Sentiment and Topics for a Product Event on Twitter</i> ” (Wedel et al., 2022)	Analisis sentimen terkait iPhone 12 pada <i>tweets</i> berbahasa Jerman dan Inggris. Akurasi pelabelan sentimen menggunakan VADER lebih tinggi (0,57) dibanding SentiWordNet (0,36).	
4.	“ <i>Attitudes Evaluation Toward COVID-19 Pandemic: An Application of Twitter</i> ”	1,4 juta <i>tweets</i> terkait sentimen <i>COVID-19</i> berhasil dianalisis menggunakan pelabelan sentimen VADER dan	

No.	Judul	Hasil	Korelasi
	<i>Sentiment Analysis and Latent Dirichlet Allocation</i> ” (Shurrab et al., 2021)	pemodelan topik LDA setiap bulannya. Peneliti menyarankan untuk melakukan <i>grid search optimization</i> dalam melakukan pemodelan topik menggunakan LDA agar dapat menghasilkan hasil yang lebih baik.	Mengombinasikan metode VADER dan LDA dalam menganalisis <i>tweets</i> . VADER dipilih karena memiliki akurasi yang lebih tinggi dibanding metode lain.
5.	“ <i>Mining non-functional requirements from App store Reviews</i> ” (Jha and Mahmoud, 2019)	Peneliti berhasil mengekstrak <i>Non-Functional Requirements (NFR)</i> pada ulasan beberapa aplikasi App Store menggunakan <i>term matching</i> . meliputi <i>usability, supportability, dependability, dan performance</i> untuk mencapai kepuasan pengguna. Ditemukan bahwa ulasan yang memiliki panjang teks 12 kata (tanpa <i>stopwords</i>) merupakan panjang teks yang optimal dan memiliki <i>precision</i> dan <i>recall</i> yang seimbang dalam melakukan klasifikasi.	Menggunakan <i>lexicon-based</i> dalam pelabelan sentimen dengan <i>threshold</i> panjang teks 12 kata.
6.	“Analisis Ulasan <i>Indie Video Game</i> Lokal pada Steam Menggunakan Analisis Sentimen dan Pemodelan Topik Berbasis <i>Latent Dirichlet Allocation</i> ” (Febrianta et al., 2021)	Pemodelan topik menggunakan LDA dilakukan setelah mengklasifikasi label sentimen ulasan <i>Game</i> menggunakan RapidMiner dan RStudio. Peneliti dapat mengekstrak istilah yang sering muncul pada topik-topik dominan.	Menganalisis sentimen suatu produk digital dari <i>user-generated content</i> , melakukan pemodelan topik menggunakan LDA.
7.	“ <i>Sentiment Analysis and Topic Detection of Mobile Banking Application Review</i> ” (Eksa Permana et al., 2020)	Penelitian ini melakukan analisis sentimen (NBC) terhadap ulasan beberapa <i>mobile banking</i> Indonesia dari Google Play Store dan dikombinasikan dengan pemodelan topik (LDA) untuk mengetahui topik apa saja yang didiskusikan dalam ulasan-ulasan pengguna. Dengan dilakukannya penelitian ini juga dapat mengetahui kelebihan dan kelemahan setiap aplikasi.	

No.	Judul	Hasil	Korelasi
8.	“ <i>Topic Modeling using The Latent Dirichlet Allocation Method on Wikipedia Pandemic Covid-19 Data in Indonesia</i> ” (Sari and Purnomo, 2022)	Pemodelan topik menggunakan LDA pada artikel Wikipedia terkait pandemi COVID-19 di Indonesia. Data berjumlah 6.658 kata. Dalam proses pemodelan LDA berpatok pada kata-kata yang sering muncul. Peneliti melakukan beberapa iterasi, masing-masing iterasi dengan jumlah topik yang berbeda sehingga ditemukan topik yang paling dibahas dan sesuai.	Pemodelan topik menggunakan LDA.
9.	“ <i>Trend Topic Analysis using Latent Dirichlet Allocation (LDA) Study Case: Denpasar People’s Complaints Online Website</i> ” (Destarani et al., 2019)	Dengan pemodelan topik menggunakan LDA pada data komplain warga Denpasar, 19 topik ditemukan dan topik ke-4 memiliki probabilitas tertinggi sebesar 0,055. Dari topik ke-4 didapat diperlukan perbaikan jalan secepatnya.	
10.	“ <i>Hybrid approach: naive bayes and sentiment VADER for analyzing sentiment of mobile unboxing video comments</i> ” (D, 2019)	Peneliti menggabungkan pendekatan <i>lexicon</i> VADER dengan <i>machine learning</i> Naïve Bayes dalam menganalisis dan memprediksi sentimen komentar video. VADER memberi pengaruh baik terhadap proses klasifikasi Naïve Bayes dengan akurasi 79,78% dan F1 Score 83,72%.	Analisis sentimen menggunakan VADER. VADER memberikan nilai akurasi yang baik.

Berdasarkan peninjauan penelitian terdahulu, metode VADER memiliki performa yang baik dalam analisis sentimen pada data yang berasal dari media sosial khususnya *tweets*. Sedangkan metode LDA juga terbukti efisien dalam melakukan pemodelan topik. Kombinasi kedua metode ini akan memberikan hasil yang komprehensif dalam menganalisis topik opini-opini sentimen masyarakat terkait QRIS pada Twitter. Adapun didapat informasi bahwa panjang teks optimal dalam melakukan klasifikasi yaitu ≥ 12 kata (Jha and Mahmoud, 2019).

2.2. Teori-teori Dasar

2.2.1. QRIS

QRIS (*Quick Response Code Indonesian Standard*) adalah suatu standar kode QR Nasional guna memfasilitasi pembayaran kode QR di Indonesia. QRIS diluncurkan oleh Bank Indonesia dan Asosiasi Sistem Pembayaran Indonesia (ASPI) pada 17 Agustus 2019. Dengan QRIS, seluruh aplikasi pembayaran digital dari PJSP manapun baik bank dan non-bank yang digunakan masyarakat, dapat digunakan di seluruh *merchant* (dari toko, warung, umkm, donasi, hingga parkir) berlogo QRIS. (Bank Indonesia, 2023; QRIS.id, 2023).

2.2.2. Text Mining

Text mining adalah sebuah teknik atau pendekatan algoritma berbasis komputer untuk mendapatkan suatu pengetahuan baru dari *pattern* tersembunyi sekumpulan teks (Hakim, 2018; Priyanto and Ma'arif, 2018). *Text mining* merupakan bidang interdisipliner yang mengambil unsur-unsur dari *information retrieval*, *data mining*, *machine learning*, statistik, dan komputasi linguistik (Attabi et al., 2018). *Text mining* bertujuan utama untuk mendapatkan informasi berkualitas tinggi dari teks. Umumnya dilakukan melalui penemuan *pattern* dan tren dengan cara seperti *statistical pattern learning*, *topic modeling*, dan *statistical language modeling*. *Text mining* berupa *task* yang umum meliputi *text categorization*, *text clustering*, *concept/entity extraction*, *production of granular taxonomies*, *sentiment analysis*, *document summarization*, dan *entity-relation modeling* (Han et al., 2012; Saraswati, 2013).

Dalam mengekstrak informasi, berikut tahapan-tahapan yang umum dilakukan pada implementasi *text mining*:

A. *Data Collection*

Data collection ialah suatu proses pengumpulan dan pengambilan data dari sumber-sumber data. *Data collection* dapat dilakukan secara berkala (per hari, per jam, atau per sekian menit) atau secara *real-time* (M.Sc, 2022). Salah satu teknik *data collection* adalah *scraping*. *Scraping* adalah teknik yang kuat dan *scalable* untuk memperoleh konten. Infrastruktur Python mendukung proses *scraping* dengan baik (Albrecht et al., 2020). *Scraping* umumnya digunakan dalam mendapatkan data yang berasal dari media sosial seperti Twitter.

B. *Data Cleaning*

Data sering mengandung *noise, errors, exceptions* atau tidak lengkap. *Noise* dan *errors* dapat membingungkan proses analisis, yang mengarah pada penemuan pola yang salah (Han et al., 2012). *Data cleaning* adalah proses menghilangkan *noise* pada data (Wettayaprasit et al., 2007). Pada tahap ini data yang tidak perlu atau tidak relevan akan dihapus karena akan mempengaruhi pada proses analisis. Biasanya digunakan dalam menyaring komentar atau *tweets* spam yang tidak ada keterkaitan dengan penelitian (Rodrigues et al., 2022).

C. *Data Preprocessing*

Data preprocessing merupakan teknik transformasi data *raw* menjadi format yang dapat dimengerti oleh mesin (Probyto Data Science and Consulting Pvt, 2020). *Data preprocessing* adalah bagian prasyarat pada domain NLP. *Data preprocessing* terdiri dari berbagai teknik yang dilakukan sesuai kebutuhan data. Teknik yang lazim dilakukan di antaranya menghapus tag HTML, *hashtag*, URL, tanda baca, dan *whitespace* (Vasant et al., 2022). Karena dataset yang digunakan dalam penelitian ini berupa *tweets*, berikut merupakan teknik *data preprocessing* yang digunakan dalam penelitian ini:

1. *Remove Username, Hashtag, & URLs*. Teknik ini digunakan untuk menghapus bagian *tweets* yang mengandung *username, hashtag, dan link* URL seperti *link* gambar atau berita, dll.
2. *Remove Duplicates*. Teknik ini dilakukan dengan *drop tweets* yang bersifat duplikat. Tidak jarang ditemukan *tweets* spam pada dataset, maka teknik ini juga akan menghapus *tweets* spam yang dapat mengoptimalkan kinerja mesin (Sarkar, 2016).
3. *Remove Punctuation* merupakan teknik penghapusan tanda baca seperti `'!\"#$%&\'()*+,-./:;<=>?@[\\]^_`{|}~'` yang terkandung dalam *tweets* (Albrecht et al., 2020).
4. *Case Folding* adalah proses mengubah semua karakter kata menjadi huruf kecil atau huruf besar (huruf kecil lebih umum) (Das and Cakmak, 2018). Ketika dua atau lebih kata dengan arti yang sama ditulis dengan huruf kapital yang berbeda, jumlah kata unik bertambah sehingga tidak efisien, oleh karena itu, lebih baik mengubahnya menjadi satu bentuk (Žižka et al., 2019).

5. *Tokenization* adalah proses memecah atau membagi data tekstual menjadi komponen yang lebih kecil yang disebut token. Token adalah komponen tekstual independen (Sarkar, 2016).
6. *Remove Stopwords*. Kata-kata yang paling sering muncul dalam teks adalah kata-kata umum seperti kata kerja bantu, kata ganti, kata keterangan, dan sebagainya. Kata-kata ini disebut *stopwords*. *Stopwords* biasanya tidak membawa banyak informasi tetapi menyembunyikan konten yang menarik karena frekuensinya yang tinggi. Oleh karena itu, *stopwords* lazim dihapuskan sebelum analisis data atau pelatihan model (Albrecht et al., 2020).
7. *Stemming* adalah proses mereduksi kata menjadi bentuk dasar (akar kata). Ini sangat berguna untuk *search engines*. Ketika pengguna ingin mencari buku tentang kewarganegaraan, jika kita menerapkan algoritma *stemming* ke sebuah kata, kemungkinan besar dokumen dengan bentuk kata lain seperti warga, negara, dll. akan muncul juga dan cocok dengan pencarian (Zhai and Massung, 2016).

D. *Feature Extraction*

1. *Bag of Words*

Bag of Words adalah salah satu teknik yang paling sederhana namun paling kuat untuk mengekstrak fitur dari dokumen teks. Inti dari model ini adalah mengubah dokumen teks menjadi vektor. Jadi setiap dokumen dikonversi menjadi vektor yang mewakili frekuensi semua kata yang berbeda yang terdapat dalam ruang vektor dokumen tertentu untuk dokumen tersebut (Sarkar, 2016).

2. TF-IDF

Bag of Words bagus untuk digunakan dalam melakukan *feature extraction*, namun *Bag of Words* sepenuhnya didasarkan pada frekuensi absolut dari kejadian kata. Hal ini memiliki beberapa masalah di mana kata yang cenderung terjadi banyak di seluruh dokumen dalam korpus, akan memiliki frekuensi yang lebih tinggi dan cenderung mengalahkan kata lain yang mungkin tidak sering muncul, tetapi lebih menarik dan efektif dijadikan sebagai fitur untuk mengidentifikasi kategori spesifik untuk dokumen. Berbeda dengan TF-IDF, TF-IDF adalah proses transformasi data dari data tekstual ke dalam numerik agar dilakukan pembobotan

pada setiap kata (Hartini et al., 2022). TF-IDF berarti *Term Frequency-Inverse Document Frequency*, kombinasi dua metrik yaitu *term frequency* dan *inverse document frequency* (Sarkar, 2016).

E. *Data Validation*

Data validation atau validasi data adalah suatu proses untuk menentukan validitas data, kelengkapan dan konsistensi data, dan memvalidasi data dapat dipercaya, akurat dan bermakna (Xie et al., 2017). Validasi data merupakan langkah penting untuk meningkatkan kualitas data. Seperti yang kita ketahui, kualitas data yang buruk akan berdampak signifikan pada pengambilan keputusan.

Validasi data perlu dilakukan guna mengetahui performa VADER dalam melabeli dan mengkategorikan *tweets* dengan tepat. Proses validasi dapat dilakukan dengan mengambil sampel secara random (*random sampling*) dari populasi atau total *tweets* yang akan digunakan dalam analisis sentimen (Rachmat and Lukito, 2016; Sari and Wibowo, 2019; Septian et al., 2019; Zuhri and Alamsyah, 2017). Adapun rumus pengambilan sampel untuk populasi yang telah diketahui adalah sebagai berikut:

$$\text{Sampel} = \frac{\left(\frac{z^2 \times p(1-p)}{e^2}\right)}{1 + \left(\frac{z^2 \times p(1-p)}{e^2 N}\right)} \quad (1)$$

Dengan:

N : Ukuran populasi

e : *Margin of error*

z : *z-score*

Dalam penelitian ini, validasi data yang dihasilkan dari proses pelabelan dilakukan secara manual dengan bantuan partisipasi manusia. Partisipasi manusia dalam validasi data sangatlah penting dalam studi ini, dikarenakan kekompleksan dari *tweets* (menggunakan nuansa sosial, seperti slang atau sarkasme) kurang mudah dikenali oleh sistem berbasis algoritma (Lappeman et al., 2020).

F. *Evaluation*

Confusion matrix atau juga yang dikenal sebagai *error matrix* merupakan alat yang sudah teruji dalam *machine learning* untuk mengevaluasi akurasi sebuah klasifikator (Düntsche and Gediga, 2020). *Confusion matrix* berbentuk tabel

kontingensi yang dapat menampilkan perbedaan antara kelas asli dan yang diprediksi untuk sekumpulan contoh yang dilabeli (klasifikasi) (Bradley, 1997).

Dalam Tabel 2.2 terdapat beberapa baris dan kolom dalam *confusion matrix*. Baris matriks *predicted value* merepresentasikan sebuah *instance* dalam nilai yang diprediksi, sementara kolom *actual value* merepresentasikan nilai aktual, atau sebaliknya. Terdapat empat sel pada matriks *output* dari *predicted value* dan *actual value*. *True Positive* (t_P), *True Negative* (t_N), *False Positive* (f_P), dan *False Negative* (f_N). t_P berarti nilai aktual dan nilai yang diprediksi sama-sama positif, t_N berarti nilai aktual positif namun nilai yang diprediksi oleh model negatif, f_P berarti nilai aktual negatif namun nilai yang diprediksi oleh model positif, dan f_N berarti keduanya nilai aktual dan yang diprediksi negatif (Das et al., 2022).

Tabel 2.2 *Confusion Matrix*

<i>Confusion Matrix</i>		<i>Actual Value</i>	
		<i>Positive</i>	<i>Negative</i>
<i>Predicted Value</i>	<i>Positive</i>	t_P	f_P
	<i>Negative</i>	f_N	t_N

Berdasarkan empat *output* pada *confusion matrix*, didapatkan data yang digunakan dalam mengevaluasi klasifikasi atau pelabelan. Berikut beberapa perhitungan yang sering digunakan beserta penjelasannya:

1. *Accuracy*. Merepresentasikan rasio dari prediksi yang benar dibandingkan dengan jumlah total prediksi. Biasa digunakan dalam mengukur seberapa baik sistem klasifikasi dalam mengidentifikasi item yang benar secara keseluruhan.

$$Accuracy = \frac{(t_P + t_N)}{(t_P + f_P + t_N + f_N)} \quad (2)$$

2. *Recall*. Rasio dari jumlah prediksi benar positif (*True Positive*) dengan jumlah total item yang sebenarnya positif. *Recall* mengukur seberapa baik sistem klasifikasi dalam menemukan semua item yang sebenarnya positif.

$$Recall = \frac{t_P}{(t_P + f_N)} \quad (3)$$

3. *Precision*. Rasio dari jumlah prediksi benar positif (*True Positive*) dengan jumlah total prediksi positif (*True Positive* dan *False Negative*). *Precision*

mengukur seberapa baik klasifikasi sistem dalam mengidentifikasi posisi yang benar.

$$Precision = \frac{t_p}{(t_p + f_p)} \quad (4)$$

4. *F1 Score*. *F1 Score* atau *F measure* juga lazim digunakan dalam mengukur akurasi. Dikalkulasikan dengan rata-rata terbobot dari *precision* dan *recall* dengan rentang dari yang terbaik 1 hingga 0 terburuk.

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (5)$$

2.2.3. Analisis Sentimen

Analisis sentimen atau juga dapat disebut *opinion mining* merupakan suatu bidang studi yang mengkaji analisis sentimen, opini, perasaan, penilaian, emosi dan sikap seseorang terhadap suatu entitas seperti permasalahan, layanan, produk, individu, organisasi, acara, dan topik dari bahasa tertulis. Analisis sentimen termasuk salah satu bidang penelitian paling aktif dalam *Natural Language Processing* (NLP) dan juga dipelajari secara luas dalam *data mining*, *web mining*, dan *text mining*. Analisis sentimen dapat diimplementasikan pada tiga level yakni level dokumen, kalimat, dan entiti atau aspek (Liu, 2012). Pada penelitian ini akan terletak pada level dokumen.

Analisis sentimen secara umum dapat dilakukan dengan menggunakan pendekatan *supervised machine learning* atau pendekatan *rule-based* berbasis *lexicon* (Albrecht et al., 2020). Indikator paling penting dalam melakukan analisis sentimen adalah kata-kata sentimen, juga disebut kata-kata opini, yakni kata-kata yang biasa digunakan untuk mengekspresikan sentimen positif atau negatif. Sebagai contoh, baik, luar biasa, dan menakjubkan merupakan kata-kata sentimen positif. Sedangkan buruk dan mengerikan termasuk kata-kata sentimen negatif. Selain itu terdapat juga frasa dan idiom, salah satu contohnya “tebal telinga”. Kata-kata sentimen dan frasa sangat penting dalam analisis sentimen karena alasan yang jelas. Daftar kata-kata dan frasa seperti itu disebut *lexicon sentiment*. Selama bertahun-tahun, peneliti telah merancang berbagai algoritma untuk menyusun *lexicon* (Liu, 2012).

A. *Valence Aware Dictionary for sEntiment Reasoning (VADER)*

Salah satu metode *lexicon-based* yang populer dan telah banyak digunakan dalam penelitian adalah VADER. VADER merupakan alat analisis sentimen yang digunakan untuk mengevaluasi teks dan menentukan sentimen positif, negatif, atau netral (Hutto and Gilbert, 2014). Algoritma VADER menggunakan kombinasi dari pendekatan *lexicon-based* dan *rule-based* untuk mengevaluasi sentimen teks. Ia menggunakan kamus kata-kata yang diklasifikasikan sebagai positif, negatif, atau netral, serta aturan sintaksis untuk menentukan konteks dari kata-kata tersebut dalam teks. Dalam kamus *lexicon* VADER disertai juga dengan bobot masing-masing kata dari rentang -4 hingga +4. Setelah menganalisis teks, VADER akan menghasilkan skor *compound*. Skor *compound* didapat dari penjumlahan skor *valence* dari setiap kata dalam *lexicon* yang lalu dinormalisasi berkisar antara -1 (sangat negatif) hingga +1 (sangat positif) (Hutto and Gilbert, 2014). Dalam melakukan klasifikasi label sentimen, *threshold* yang umum digunakan adalah sebagai berikut:

$$\begin{cases} \text{skor compound} \leq -0.05 = \text{negatif} \\ -0.05 < \text{skor compound} < 0.05 = \text{netral} \\ \text{skor compound} \geq 0.05 = \text{positif} \end{cases} \quad (6)$$

Dalam melakukan analisis sentimen menggunakan dataset *tweets*, VADER terbukti lebih unggul memiliki akurasi tertinggi sebesar 0,96 dibanding 12 metode lainnya sebagaimana ditunjukkan pada Tabel 2.3.

Tabel 2.3 Perbandingan F1 Score VADER (Hutto and Gilbert, 2014)

Metode	Akurasi Kelas Klasifikasi menggunakan F1 Score			
	<i>Tweets</i>	Ulasan Film	Ulasan Amazon	NYT Editorial
VADER	0,96	0,61	0,63	0,55
Naïve Bayes (<i>Tweets</i>)	0,84	0,53	0,53	0,42
<i>Maximum Entropy (Tweets)</i>	0,83	0,56	0,58	0,45
<i>SVM Classification (Tweets)</i>	0,83	0,56	0,55	0,46
<i>SVM Regression (Tweets)</i>	0,65	0,49	0,51	0,46

Naïve Bayes (Ulasan Film)	0,56	0,75	0,49	0,44
<i>Maximum Entropy</i> (Ulasan Film)	0,56	0,75	0,51	0,45
Naïve Bayes (Ulasan Amazon)	0,69	0,55	0,61	0,48
<i>Maximum Entropy</i> (Ulasan Amazon)	0,67	0,55	0,60	0,43
<i>SVM Classification</i> (Ulasan Amazon)	0,64	0,55	0,58	0,42
<i>SVM Regression</i> (Ulasan Amazon)	0,54	0,49	0,48	0,44
Naïve Bayes (NYT Editorial)	0,59	0,56	0,51	0,49
<i>Maximum Entropy</i> (NYT Editorial)	0,58	0,55	0,51	0,50

B. *InSet Lexicon*

Berdasarkan penelitian serupa, ntlk VADER dimanfaatkan namun dimodifikasi dengan *custom lexicon*. Hal ini dikarenakan *dictionary default* VADER berisikan *lexicon* berbahasa Inggris sehingga harus dimodifikasi dengan *custom lexicon* yang cocok dengan bahasa yang digunakan dalam dataset (Ekaputri and Akbar, 2022; Rintyarna, 2021). *Lexicon* Bahasa Indonesia yang akan digunakan adalah *InSet Lexicon*, terdiri dari kumpulan 3.609 kata positif dan 6.609 kata negatif. *InSet Lexicon* dipilih untuk penelitian ini karena berdasarkan *literature review* (Fauziah et al., 2021), *InSet Lexicon* memiliki nilai akurasi yang lebih tinggi dibanding SentiWordNet. *InSet Lexicon* dikembangkan dari dataset *tweets* juga sehingga sangat cocok untuk penelitian ini (Koto and Rahmaningtyas, 2017). Setiap kata-kata pada *InSet Lexicon* telah disertai dengan bobot dari rentang -5 hingga +5. Cuplikan daftar kata yang tersedia dapat dilihat pada Tabel 2.4.

Meskipun *lexicon* yang tersedia pada *InSet* hanyalah positif dan negatif, klasifikasi pelabelan sentimen akan menghasilkan tiga label yaitu positif, negatif, dan netral (Fadhillah et al., 2019; Firdaus et al., 2021; Musfiroh et al., 2021; Nurkasanah and Hayaty, 2022). Dalam melakukan klasifikasi label sentimen, *threshold* yang akan digunakan adalah sebagai berikut:

$$\begin{cases} \text{skor compound} < 1 = \text{negatif} \\ \text{skor compound} = 0 = \text{netral} \\ \text{skor compound} > 1 = \text{positif} \end{cases} \quad (7)$$

Tabel 2.4 Cuplikan Daftar Kata pada InSet *Lexicon*

Kata Negatif	Bobot	Kata Positif	Bobot
berkabung	-5	bersyukur	5
pusing	-4	sukarela	4
keluar	-3	harap	3
licin	-2	meringankan	2
dahulu	-1	ujian	1

2.2.4. Pemodelan Topik

Pemodelan topik atau *topic modeling* memiliki peran penting dalam ilmu komputer untuk *text mining* dan *Natural Language Processing* (NLP). Pemodelan topik diterapkan secara luas dalam NLP untuk penemuan topik dan *semantic mining* dari dokumen yang tidak berurutan (Blei et al., 2003). Pemodelan topik adalah suatu teknik yang digunakan untuk menemukan struktur tersembunyi dari kumpulan dokumen (Barde and Bainwad, 2017). Pemodelan topik dilakukan dengan membuat mesin membaca sejumlah besar teks dan mencari kata-kata yang sering muncul bersama-sama. Kemudian, mesin akan mengelompokkan kata-kata tersebut ke dalam beberapa kelompok yang disebut “topik”. Dalam pemodelan topik, topik adalah daftar kata yang muncul dalam metode yang signifikan secara statistik.

Contohnya, terdapat beberapa dokumen dengan lebih dari satu topik. Mesin akan membaca semua dokumen tersebut dan mencari kata-kata yang sering muncul bersama-sama, seperti “gen”, “dna”, “genetik”, “data”, “nomor”, dan “komputer”. Kemudian mesin akan mengelompokkan kata-kata tersebut ke dalam masing-masing topik misalnya “genetik” dan “komputasi” (Blei et al., 2003). Pemodelan topik menemukan topik-topik dalam teks dengan mencari kata-kata yang sering muncul bersama-sama dan mengelompokkannya ke dalam topik yang berbeda. Hal ini sangat berguna dalam mengekstrak informasi dari teks yang sangat besar dan membuatnya lebih mudah untuk dianalisis dan dipahami oleh manusia.

Pemodelan topik tidak dapat memahami makna dan konsep kata dalam dokumen teks (Schmiedel et al., 2018). Sebaliknya, mereka menganggap bahwa

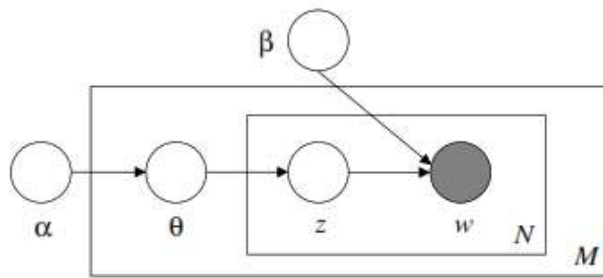
bagian manapun dari teks tersebut dikombinasikan dengan memilih kata dari kumpulan kata yang memungkinkan menjadi topik, di mana setiap kumpulan kata tersebut sesuai dengan topik. Proses ini dilakukan secara berulang-ulang hingga distribusi kata yang paling mungkin dijadikan topik ditemukan. Pemodelan topik dapat diimplementasikan pada teks berupa email, bab buku, blog *posts*, artikel jurnal, *tweets*, dan jenis teks tidak terstruktur lainnya.

Pemodelan topik merupakan teknik yang sangat kuat dalam *text mining* untuk *data mining*, *latent data discovery*, dan mencari hubungan antara data dan dokumen teks. Telah banyak studi literatur yang mengkaji pemodelan topik dan diterapkan dalam berbagai bidang seperti *software engineering*, ilmu medis, ilmu politik, dll. Terdapat berbagai metode untuk pemodelan topik, salah satu metode yang paling populer dan sering digunakan dalam berbagai studi adalah *Latent Dirichlet Allocation* (LDA).

A. *Latent Dirichlet Allocation* (LDA)

Latent Dirichlet Allocation (LDA) adalah sebuah teknik pemodelan probabilistik yang digunakan untuk mengkategorikan atau menemukan topik-topik dalam teks (Blei et al., 2003). Dalam metode ini, dokumen adalah objek yang dapat diamati, sementara topik, distribusi topik per-dokumen, dan pembagian setiap kata pada topik per-dokumen adalah struktur tersembunyi. Oleh karena itu, algoritma ini disebut LDA (Putra, 2017).

LDA didasarkan dengan asumsi bahwa kata-kata dihasilkan oleh topik dengan distribusi bersyarat tetap (Blei et al., 2003). Jadi LDA menggunakan variabel *latent* yang tidak diketahui pasti topik apa yang mewakili setiap teks atau kata, namun LDA dapat menentukan probabilitas setiap teks yang diterapkan pada setiap topik dan menentukan probabilitas setiap topik yang memiliki kata-kata tertentu. Terdapat notasi *plate* yang dapat secara ringkas merepresentasikan secara visual ketergantungan di antara parameter model. Notasi *plate* dapat dilihat pada Gambar 2.1.



Gambar 2.1 Notasi *Plate* LDA (Blei et al., 2003)

Dengan:

- M : Jumlah dokumen
- N : Jumlah kata dalam setiap dokumen
- α : Distribusi topik setiap dokumen
- β : Distribusi kata setiap topik
- z : Topik untuk kata ke- n pada dokumen ke- m
- w : Kata spesifik
- θ : Distribusi topik untuk dokumen ke- m

Terdapat dua persegi panjang dan beberapa variabel di dalam dan luarnya. Persegi panjang terluar melambangkan jumlah dokumen dalam korpus yang hendak dianalisis (M). Persegi panjang di dalam melambangkan jumlah kata dalam setiap dokumen (N).

Pada tingkatan korpus (dokumen), terdapat dua *dirichlet prior* parameter yaitu *alpha* (α) dan *beta* (β). *Alpha* (α) merupakan parameter yang digunakan untuk menghitung seberapa banyak topik yang dibahas dalam suatu dokumen. Jika nilai *alpha* semakin tinggi, menandakan bahwa dokumen tersebut membahas lebih banyak topik (campuran), sedangkan jika semakin rendah artinya dokumen hanya membahas beberapa topik tertentu. *Beta* (β) adalah parameter yang digunakan untuk menghitung distribusi kata dalam suatu topik. Jika nilai *beta* semakin tinggi, maka topik tersebut mengandung lebih banyak kata dan jika sebaliknya, maka topik tersebut hanya mengandung beberapa kata tertentu (lebih spesifik).

Dalam tingkatan dokumen, variabel θ mewakili distribusi suatu topik dalam dokumen tertentu. Jika nilai θ semakin tinggi, dokumen tersebut membahas lebih banyak topik, sedangkan jika sebaliknya maka dokumen tersebut lebih spesifik dalam membahas topik tertentu. Variabel z dan w adalah variabel yang berada pada

tingkat kata. Variabel z mewakili sebuah topik dari kata tertentu dalam suatu dokumen. Variabel w mewakili suatu kata yang berkaitan dengan topik tertentu (lebih spesifik) yang terdapat dalam sebuah dokumen (Agustina, 2017; Blei et al., 2003; Putra, 2017).

LDA merupakan model Bayesian yang memformulasi ulang pLSA dengan mengganti variabel indeks menjadi parameter acak berupa θ (parameter multinomial) (Blei et al., 2003). Seperti yang telah dijelaskan pada Gambar 2.1, distribusi θ dipengaruhi oleh α dan hubungan antara variabel z dan w bergantung pada β . Berikut merupakan rumus probabilitas himpunan seluruh kata yang diobservasi oleh w :

$$p(w|\alpha, \beta) = \int p(\theta|\alpha) \left[\prod_{n=1}^N \sum_{z_n} p(z_n|\theta) p(w_n|z_n, \beta) \right] d\theta \quad (8)$$

Dengan:

- $p(\theta|\alpha)$: Distribusi *dirichlect* dokumen berdasarkan topik
- $p(z_n|\theta)$: Probabilitas munculnya topik dalam suatu dokumen
- $p(w_n|z_n, \beta)$: Probabilitas munculnya kata dalam suatu topik

B. Koherensi Topik

Perkembangan pemodelan topik tidak terlepas dari tantangan dalam seberapa mudah topik-topik yang dihasilkan dapat diinterpretasi oleh manusia (Aletras and Stevenson, 2013). Koherensi topik telah diusulkan sebagai metode evaluasi untuk pemodelan topik (Newman et al., 2010b, 2010a). Koherensi topik didefinisikan sebagai rata-rata keterkaitan semantik antar kata-kata topik, dan menghasilkan topik yang lebih mudah diinterpretasi oleh manusia (Newman et al., 2010a). Pengukuran koherensi terbaik dihasilkan dengan kombinasi C_v , *indirect cosine* dengan NPMI, dan *boolean sliding window* (Röder et al., 2015).

C_v mengukur kesamaan antara kata-kata dalam satu topik. Koherensi C_v didasarkan pada *pairwise similarity* antara kata-kata dalam topik. *Pairwise similarity* adalah metode yang digunakan untuk mengukur sejauh mana dua objek atau entitas mirip satu sama lain. Dalam konteks koherensi C_v , *pairwise similarity* mengacu pada sejauh mana kata-kata dalam topik terkait satu sama lain berdasarkan

kemunculan bersama mereka dalam teks. Dengan mengukur *pairwise similarity* antara kata-kata dalam topik, kita dapat menilai sejauh mana topik tersebut koheren dan mudah diinterpretasikan. Semakin tinggi nilai C_v dari sebuah topik, maka topik tersebut dianggap semakin koheren atau teratur dalam menyatakan konsep (Mifrah and Benlahmar, 2020).

Adapun metrik pengukuran koherensi lain yang umum digunakan yaitu $C_{U\ Mass}$. $C_{U\ Mass}$ bertujuan untuk mengukur seberapa sering kata-kata yang muncul dalam pemodelan topik dan keterkaitan satu sama lain secara semantik. Metrik ini menghitung probabilitas log dari kemunculan kata-kata yang saling terkait dalam suatu konteks, dan digunakan untuk mengoptimalkan kualitas koherensi semantik dari pemodelan topik. Jika nilai $C_{U\ Mass}$ semakin mendekati 0, semakin tinggi juga kualitas koherensi semantik dari model topik tersebut (Mimno et al., 2011). Salah satu kelebihan dari penggunaan metrik ini adalah waktu yang dibutuhkan lebih cepat dari metrik lain (Röder et al., 2015).

2.2.5. *Library* yang digunakan

Dalam menunjang pemrosesan data dan analisis penelitian, berikut *library-library* yang digunakan:

- A. Pandas adalah sebuah *open source library* Python spesialis untuk analisis data. Pandas dikembangkan berdasarkan dari kebutuhan untuk analisis data yang menyediakan semua alat bantu untuk pengolahan data, ekstraksi data, dan manipulasi data. *Library* ini dirancang berdasarkan perpustakaan NumPy (McKinney, 2010; Nelli, 2018).
- B. Regex (*Regular Expression*) adalah sebuah ekspresi atau notasi khusus yang digunakan untuk menentukan pola pada sekumpulan teks atau *string* (Aho, 1990). Regex sering digunakan dalam bahasa pemrograman, seperti Python, untuk mencocokkan, mencari, dan memanipulasi *string* sesuai dengan pola tertentu. Terdapat *library* Python bawaan bernama *re*, menyediakan berbagai fungsi dalam pengoperasian regex. Modul *re* sangat berguna dalam manipulasi *string* yang rumit pada Python (Nelli, 2018). Pada penelitian ini, *library re* dimanfaatkan pada tahap *data preprocessing* dalam menghapus *username*, *hashtag*, dan URL pada *tweets* juga *tokenization*.

- C. Sastrawi adalah sebuah *library* yang berguna dalam melakukan proses *stemming* dan penghapusan *stop words* pada teks berbahasa Indonesia (“Sastrawi,” 2023).
- D. NLTK (*Natural Language Toolkit*) adalah sebuah *library* pada bahasa pemrograman Python yang digunakan untuk pengolahan bahasa alami (*Natural Language Processing*). NLTK menyediakan berbagai alat bantu dan sumber daya untuk mengekstrak informasi, menganalisis, dan memanipulasi teks bahasa alami (Bird et al., 2009).
- E. Gensim adalah sebuah *open-source library* pada bahasa pemrograman Python yang digunakan untuk pengolahan teks besar dan *topic modeling*. Gensim menyediakan berbagai alat untuk memproses dan menganalisis teks besar, seperti representasi vektor kata, *topic modeling* (seperti *Latent Dirichlet Allocation*), dan algoritma Word2Vec untuk mempelajari representasi vektor kata (Řehůřek and Sojka, 2010).
- F. Matplotlib adalah *library* Python yang memfokuskan pada pengembangan grafik 2D dan 3D. Di antara semua fitur yang membuatnya menjadi alat paling banyak digunakan dalam representasi grafis data, ada beberapa hal yang menonjol, yaitu kemudahan dalam penggunaan, *gradual development* dan visualisasi data yang interaktif, perkembangan bertahap dan visualisasi data interaktif, *expressions* dan teks dalam LaTeX, kontrol yang lebih besar atas elemen grafis, ekspor dalam banyak format seperti PNG, PDF, SVG, dan EPS (Hunter, 2007; Nelli, 2018).

2.3. Integrasi Keilmuan

Selama pandemi *COVID-19*, masyarakat Indonesia lebih berhati-hati dalam menjaga kesehatan salah satunya menghindari kontak fisik dan meminimalisir penggunaan uang kertas. Adanya QRIS dalam mengintegrasikan berbagai metode pembayaran digital merupakan sebuah *game changer* bagi masyarakat Indonesia yang sebelumnya lebih sering bertransaksi menggunakan uang kertas. Ternyata setelah pandemi usai-pun penggunaan pembayaran digital semakin meningkat (Rusnawati et al., 2022). QRIS hadir memudahkan transaksi jual beli tidak hanya bagi pelanggan namun juga para *merchant* dalam mengorganisir keuangan.

Dampak positif besar yang diberi QRIS pada kehidupan sehari-hari masyarakat Indonesia menjadi salah satu dorongan dilakukannya penelitian ini. Selain itu, salah satu tujuan penelitian analisis sentimen berbasis topik terkait QRIS adalah membantu dalam mengambil keputusan (*data-driven decision making*) dalam pengembangan layanan QRIS agar dapat lebih baik dan tepat sesuai dengan kebutuhan masyarakat. Berdasarkan hasil wawancara dengan Ibu Wiwin Luqna Hunaida, M.Pd.I, Dosen Sistem Informasi UINSA yang mengampu mata kuliah Studi Al-Qur'an dan Studi Hadits, terdapat suatu potongan hadis yang berkaitan dengan penelitian ini:

...أَنْتُمْ أَعْلَمُ بِأَمْرِ دُنْيَاكُمْ

Terjemah:

“...Engkau sekalian lebih mengetahui tentang urusan dunia.” (HR. Muslim: 2363)

Dari potongan hadits di atas, Narasumber menjelaskan bahwa terdapat korelasi dengan kemajuan teknologi yang berkembang pesat dan pentingnya menuntut ilmu dan mengimplementasikan ilmu pengetahuan sains dan teknologi dalam menciptakan kebaikan-kebaikan di bumi. Terdapat juga ayat Al-Qur'an yang bersinggungan terkait bagaimana kita harus melakukan kebaikan, memberi manfaat, dan memudahkan urusan orang lain. Surah Al-Qashash ayat 77 berbunyi:

وَأَبْنَعْ فِيمَا ءَاتَاكَ اللَّهُ الدَّارَ الْآخِرَةَ ۖ وَلَا تَنْسَ نَصِيبَكَ مِنَ الدُّنْيَا ۖ وَأَحْسِنَ كَمَا
أَحْسَنَ اللَّهُ إِلَيْكَ ۖ وَلَا تَبْغِ الْفُسَادَ فِي الْأَرْضِ ۖ إِنَّ اللَّهَ لَا يُحِبُّ الْمُفْسِدِينَ

Terjemah:

“Dan carilah (pahala) negeri akhirat dengan apa yang telah dianugerahkan Allah kepadamu, tetapi janganlah kamu lupakan bagianmu di dunia **dan berbuat baiklah (kepada orang lain) sebagaimana Allah telah berbuat baik kepadamu, dan janganlah kamu berbuat kerusakan di bumi. Sungguh Allah tidak menyukai orang yang berbuat kerusakan.**”(Al-Qashash: 77)

Dalam buku Tafsir Al-Mishbah karya M. Quraish Shihab menjelaskan tafsir ayat di atas bahwa “Jadikanlah sebagian dari kekayaan dan karunia yang Allah berikan kepadamu di jalan Allah dan amalan untuk kehidupan akhirat. Janganlah kamu cegah dirimu untuk menikmati sesuatu yang halal di dunia. Berbuat baiklah

kepada hamba-hamba Allah sebagaimana Allah berbuat baik kepadamu dengan mengaruniakan nikmat-Nya” (Shihab, 2000). Karunia yang diberikan Allah pada kita dapat berupa apapun, salah satunya yaitu ilmu. Dengan ilmu teknologi yang didapat manusia lalu dikembangkan menjadi sesuatu yang dapat memudahkan urusan orang lain, seperti dibuatnya QRIS dalam memudahkan transaksi dan dilakukannya penelitian dalam membantu pihak yang berkepentingan untuk mengambil keputusan.

M. Quraish Shihab juga menegaskan, dalam berbuat baik dan memudahkan urusan orang lain, kata *ahsin*, berasal dari kata *hasan* yang memiliki arti baik, *ahsin* di sini berbentuk perintah dan membutuhkan objek. Tetapi pada ayat ini tidak menyebutkan objek sehingga dapat meliputi segala sesuatu yang dapat disentuh oleh kebaikan, seperti kebaikan terhadap lingkungan, manusia, harta benda, tumbuh-tumbuhan, dan binatang (Shihab, 2000).

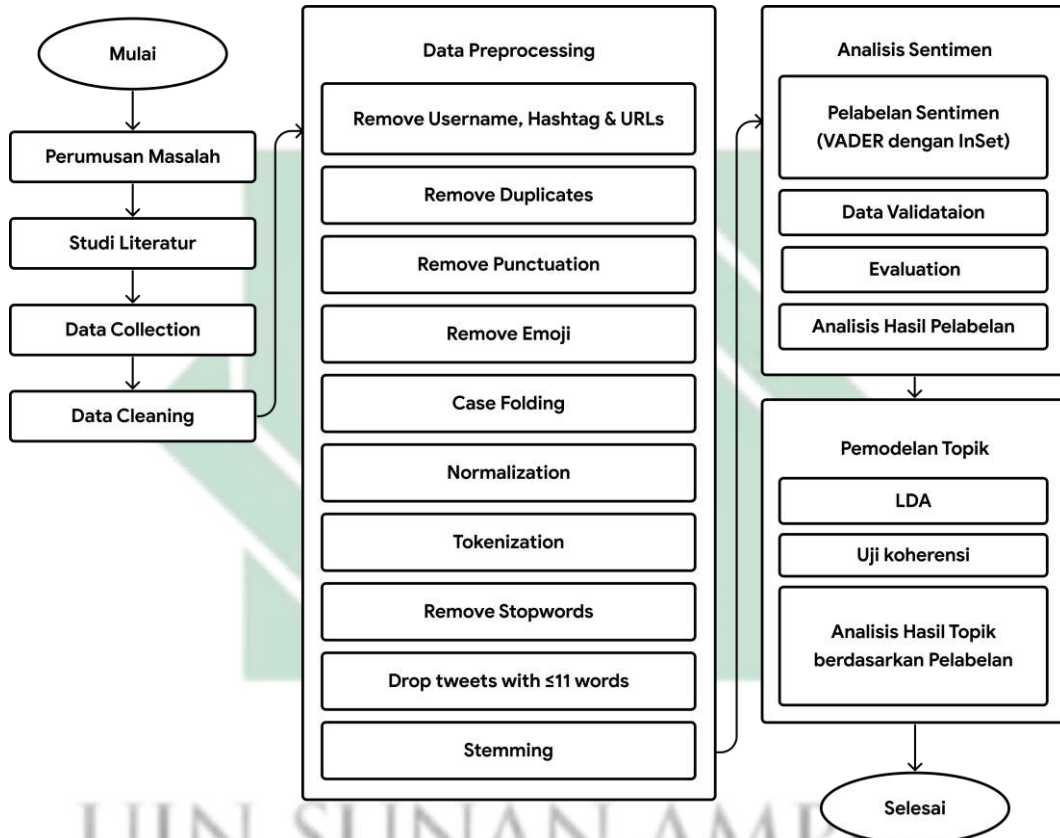


UIN SUNAN AMPEL
S U R A B A Y A

BAB III METODE PENELITIAN

3.1. Tahapan Penelitian

Agar dapat memudahkan dalam memahami proses penyusunan dan langkah-langkah penelitian dengan detail, dibuatlah diagram alir yang dapat dilihat pada Gambar 3.1.



Gambar 3.1 Diagram Alir Langkah-langkah Penelitian

Berdasarkan diagram alir di atas, berikut uraian penjelasan detail setiap langkah-langkah penelitian:

3.1.1. Perumusan Masalah

Sebuah penelitian dilakukan untuk menyelesaikan permasalahan. Perumusan masalah muncul disebabkan oleh fenomena yang melatarbelakangi penelitian sesuai pada penjelasan Latar Belakang. QRIS telah digunakan secara tersebar di seluruh penjuru Indonesia dan akan dapat digunakan juga di beberapa negara tetangga. Dalam mengevaluasi dan meningkatkan pengalaman pengguna

QRIS, salah satu cara untuk mengetahui opini masyarakat adalah dengan melakukan analisis sentimen dengan pemodelan topik.

3.1.2. Studi Literatur

Peninjauan pada penelitian terdahulu dilakukan guna menambah pengetahuan dalam memahami konsep dan topik penelitian. Studi literatur dilakukan bersumber dari jurnal dan buku. Literatur yang dipelajari meliputi penelitian-penelitian yang melakukan analisis sentimen terkait *mobile payment*, analisis sentimen menggunakan dataset *tweets*, serta implementasi VADER, LDA, dan kombinasi keduanya sebagaimana yang telah dijelaskan pada Tabel 2.1.

3.1.3. Data Collection

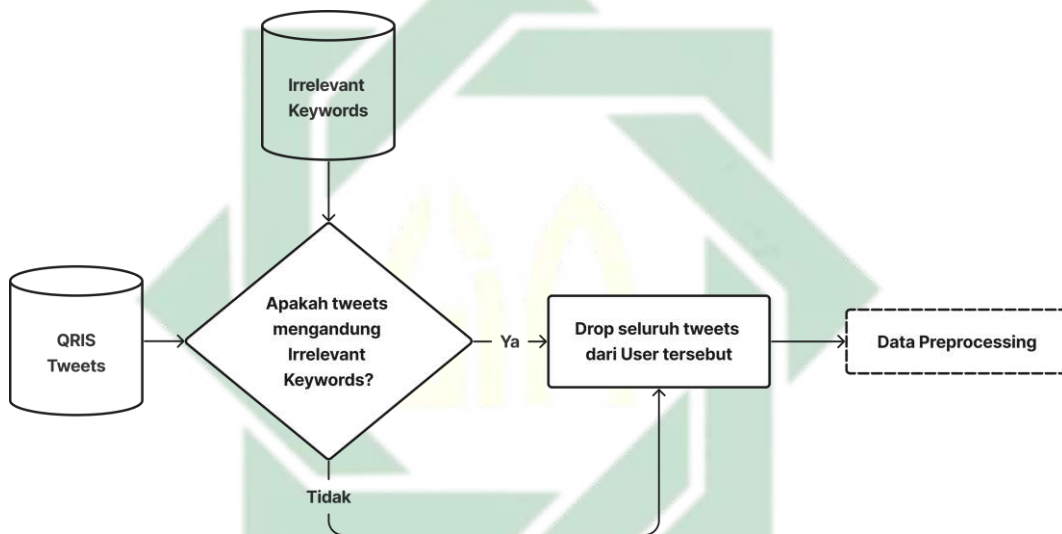
Data *tweets* didapatkan melalui proses *scraping* menggunakan *library* *snsrape* dalam bahasa pemrograman Python. Data *tweets* yang diambil merupakan *tweets* dengan kata kunci “QRIS” pada rentang tanggal sejak dirilisnya QRIS 17 Agustus 2019 hingga 17 Agustus 2022. Cuplikan dataset dapat dilihat pada Tabel 3.1.

Tabel 3.1 Dataset *Tweets* dengan *Keyword* QRIS

Tanggal	<i>Tweets</i>	Likes
2022-07-27 06:52:22+00:00	“Lupakan kripto, terobosan keuangan terbesar di dunia saat ini mungkin adalah QRIS”	15.451
2022-08-05 06:30:46+00:00	“Dari mesin ATM sampe bahas <i>fraud</i> , polisi. Mbak, sistem pembayaran kita sangat berdikari kita punya GPN, QRIS, BI FAST, klo transaksi dlm negeri kita jauh lbh canggih dan murah, dr 2018 Visa Master kalang kabut krn BI ngeluarin GPN”	10.639
2021-09-12 07:51:07+00:00	“Kayaknya masih banyak yang belum paham ya kalau segala jenis <i>e-wallet</i> di Indonesia sudah pakai QRIS -- artinya bisa bayar lintas app/bank. Kapan itu mau bayar di Aeon pake Mandiri minta QRIS kasirnya kekeuh cuman bisa Gopay doang”	3.567
2022-03-29 05:27:22+00:00	“Saya <i>crypto skeptics</i> . Dan menurut saya QRIS + Transfer Bank via VA itu revolusi. Setiap kali saya menerangkan tentang sistem pembayaran dengan <i>virtual account</i> , dan transfer antarbank yang instan dengan biaya transfer kurang lebih cuma 20 sen, ga ada yang ga <i>amazed</i> ”	3.212

3.1.4. Data Cleaning

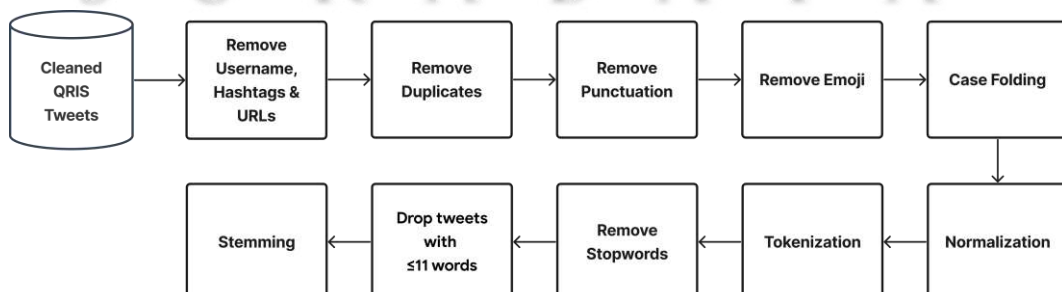
Pada tahapan ini, *tweets* yang memungkinkan berupa spam dibersihkan. Pembersihan *tweets* spam dilakukan dengan *drop* seluruh *tweets* dari *user* yang memiliki *tweets* mengandung kata dari daftar kata yang tidak relevan dengan penelitian. Kata-kata yang tidak relevan dikumpulkan secara manual. Contoh *tweets* yang tidak relevan yaitu seperti *tweets* spam penjual *subscription* produk digital yang menggunakan kata kunci QRIS untuk pembayaran. Hal ini sangat berpengaruh dikarenakan setiap detiknya, jika melakukan pencarian dengan kata kunci “QRIS” akan selalu ada *tweets* spam tersebut.



Gambar 3.2 Diagram Alir *Data Cleaning*

3.1.5. Data Preprocessing

Sebelum memproses data dalam menganalisis sentimen dan pemodelan topik, perlu dilakukan pra-proses data yang terdiri dari beberapa tahapan dalam membersihkan dan menyiapkan data agar siap diolah.



Gambar 3.3 Diagram Alir *Data Preprocessing*

Berikut penjelasan detail tahapan pra-proses data:

1. *Remove Username, Hashtag, URLs*

Tahapan ini perlu dilakukan pada *tweets* berupa balasan, *tweets* yang menyebutkan *username* atau *hashtag*, dan *tweets* yang mengandung URL *link* maupun *link* gambar. Tahapan ini terbukti dapat meningkatkan akurasi pelabelan sentimen (Keerthi Kumar and Harish, 2018).

2. *Remove Duplicates*

Menghapus *tweets* duplikat dilakukan setelah melakukan pembersihan URL. *Link* URL pada *tweets* biasanya berupa *link* gambar, *link* berita, dan sebagainya. Karena gambar yang tercantum di sebuah *tweets* berupa *link* URL yang *unique*, maka jika pada *tweets* spam terdapat *link unique* akan terdeteksi berupa *tweets* yang sama hanya saat setelah dilakukan pembersihan URL.

Tabel 3.2 Ilustrasi *Tweets* Duplikat dengan *Link* URL

Tanggal	<i>Tweets</i>
2020-03-13 10:22:13+00:00	“QRIS sudah diberlakukan di UMKM Kota Jakarta https://t.co/h0KkBWmlj ”
2020-03-13 10:23:04+00:00	“QRIS sudah diberlakukan di UMKM Kota Jakarta https://t.co/12jsGHn86 ”

3. *Remove Punctuation*

Penghapusan tanda baca juga dapat mempengaruhi akurasi (Etaiwi and Naymat, 2017). Pada tahapan ini dibantu oleh modul Python *string.punctuation* yang memudahkan dalam menghapus tanda baca di setiap *tweets*.

4. *Remove Emoji*

Emoji dapat mempengaruhi dalam proses pelabelan dan pemodelan topik (Keerthi Kumar and Harish, 2018). Bagian emoji dihapus dari *tweets* dikarenakan belum tersedia *library* yang dapat mengonversi emoji menjadi teks yang merepresentasi dalam Bahasa Indonesia.

5. *Case Folding*

Penggunaan huruf besar atau kecil dianggap tidak memiliki perbedaan. *Case folding* merupakan tahapan dalam pra-proses data yang merubah semua huruf pada *tweets* menjadi *lowercase* agar setara, sehingga ‘BAGUS’ sama dengan ‘bagus’.

6. Normalization

Indonesia merupakan negara kepulauan yang juga kaya akan kebudayaan, ras, dan suku. Dalam berbahasa sehari-hari, masyarakat Indonesia cenderung menggunakan bahasa bervariasi bercampur dengan bahasa daerah dan memakai *slang*. Dalam hal ini perlu dilakukan penanganan kata-kata di luar kosakata atau *out of vocabulary* (OOV), karena banyaknya kata dan frasa percakapan yang digunakan di media sosial. Salah satu cara dalam menangani hal tersebut adalah dengan menormalisasi kata-kata OOV berdasarkan *Colloquial Lexicon* yang dikembangkan oleh (Aliyah Salsabila et al., 2018). *Colloquial Lexicon* ini dikembangkan dari media sosial juga. Cuplikan daftar *dictionary Colloquial Lexicon* ditampilkan pada Tabel 3.3.

Tabel 3.3 *Colloquial Lexicon*

Teks	Arti	Teks	Arti
kapolda	kepala kepolisian daerah	aj	saja
adlah	adalah	aja	saja
adlh	adalah	ajaa	saja
aer	air	nyokap	ibu
ahir	akhir	brkfst	sarapan

Dapat dilihat bahwa *dictionary* ini cukup lengkap sehingga tidak hanya menormalisasi *slang* namun juga kata-kata yang lazim diketikkan dalam media sosial termasuk *elongated words* dan kata-kata yang menggunakan ejaan yang salah.

7. Tokenization

Tokenization merupakan metode untuk membagi kalimat menjadi kata-kata individu. Selain itu, dalam proses *tokenization*, juga dapat dihitung berapa kali kata tersebut muncul dalam sebuah *tweets*, yang sering disebut sebagai token. Contoh *tokenization* yaitu merubah “qris sudah diberlakukan di umkm kota jakarta” menjadi “[qris, sudah, diberlakukan, di, umkm, kota, jakarta]”.

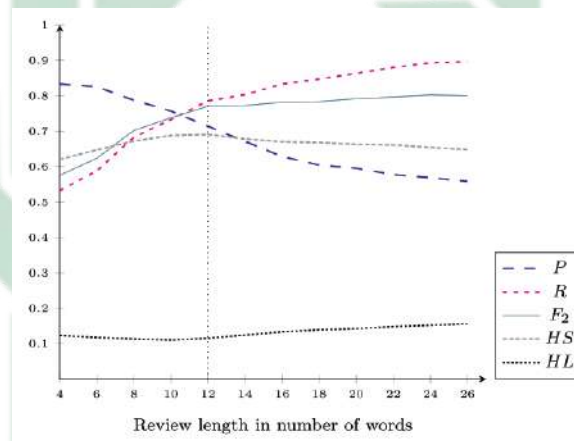
8. Remove Stopwords

Opini dalam Bahasa Indonesia banyak mengandung kata-kata yang tidak penting (*stop words*) dan kata-kata gaul (*slang words*) (Fauziah et al., 2021). Pada

tahapan ini penghapusan kata tidak penting dilakukan agar dapat meningkatkan efisiensi dalam klasifikasi pelabelan sentimen (Etaiwi and Naymat, 2017). Tahapan menghapus *stop words* berbahasa Indonesia ini memanfaatkan *library* Sastrawi (“Sastrawi,” 2023).

9. Drop Tweets

Panjang teks dalam melakukan pemodelan topik sangatlah mempengaruhi (Sbalchiero and Eder, 2020). Pada penelitian ini hanya menggunakan *tweets* yang memiliki ≥ 12 kata, sebagaimana yang dilakukan oleh (Jha and Mahmoud, 2019) dalam melakukan analisis sentimen ulasan aplikasi *App store*. Ditemukan bahwa teks dengan panjang 12 kata memiliki performa yang baik dan seimbang dalam *dictionary matching*, hal ini dilihat dari *precision* dan *recall* yang dihasilkan.



Gambar 3.4 Performa *Dictionary Matching* (Jha and Mahmoud, 2019)

10. Stemming

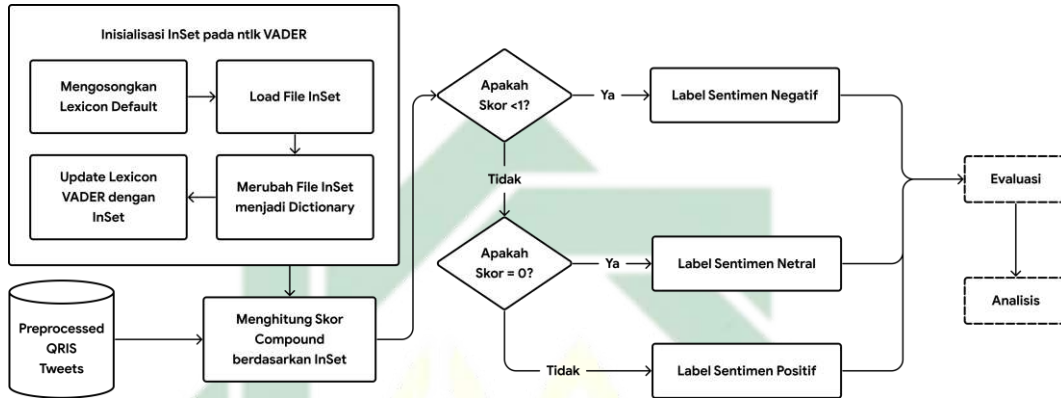
Stemming adalah proses konversi akar atau dasar kata dari kata. Proses *stemming* bahasa Indonesia ini juga dilakukan menggunakan *library* Sastrawi (“Sastrawi,” 2023). *Stemming* dapat meningkatkan efisiensi waktu yang dibutuhkan dalam pelabelan sentimen. Tabel 3.4 menampilkan contoh proses transformasi kata ke akar kata. Sastrawi mampu menangani kata-kata imbuhan jenis prefiks, sufiks, dan infiks.

Tabel 3.4 *Stemming* dengan Sastrawi (“Sastrawi,” 2023)

Kata	Akar Kata	Kata	Akar Kata
meresmikan	resmi	memperbaiki	baik
terpercaya	percaya	penjual	jual
jalan	jalan	kinerja	kerja

3.1.6. Analisis Sentimen

Klasifikasi pelabelan sentimen dilakukan menggunakan *library nltk* VADER yang *lexicon default*-nya diganti dengan InSet *Lexicon*. Proses inialisasi InSet dapat dilihat pada Gambar 3.5. Setelah InSet *Lexicon* Siap digunakan, data *tweets* yang sudah melalui tahap *preprocessing* akan dihitung skor *compound* berdasarkan bobot kata pada InSet *Lexicon*.



Gambar 3.5 Diagram Alir Klasifikasi Sentimen

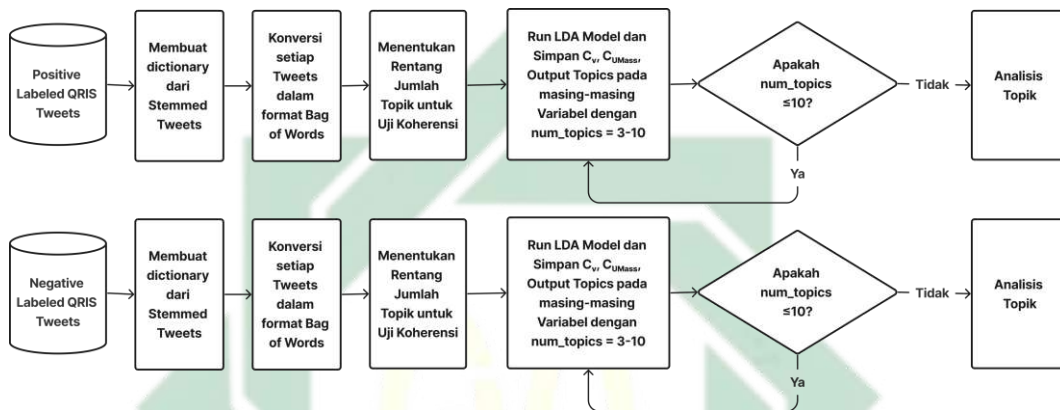
Setelah pelabelan sentimen berhasil diimplementasi menggunakan InSet *Lexicon*, validasi dan evaluasi pelabelan dilakukan dengan pengambilan data sampel menggunakan teknik *random sampling* (Anastasia and Budi, 2016; Putri and Ridwan, 2023). *Random sampling* digunakan untuk mengurangi bias dan meningkatkan representasi dari populasi yang lebih besar (Noor et al., 2022). Jumlah sampel ditentukan berdasarkan rumus (1) pad Bab II. Validasi data dilakukan secara manual dengan partisipasi manusia. Partisipan yang akan ikut serta dalam validasi pelabelan harus memiliki kualifikasi berbahasa Indonesia secara *native* dan memiliki pengetahuan seputar teknologi khususnya QRIS. Pengguna dompet digital di Indonesia didominasi oleh masyarakat yang berusia 18-24 tahun (Mediatama, 2022), sehingga menjadi target utama partisipan dalam proses validasi ini. Selanjutnya hasil validasi data akan dievaluasi menggunakan *confusion matrix*. Jika hasil evaluasi dapat diterima, maka dilanjutkan analisis sentimen dari masing-masing label.



Gambar 3.6 Diagram Alir Data Validation & Evaluation

3.1.7. Pemodelan Topik

Diagram alir proses pemodelan topik dapat dilihat pada Gambar 3.7. Setelah data *tweets* berhasil didapat pelabelan sentimennya, *feature extraction* dilakukan pada masing-masing label agar dapat diproses oleh LDA. *Tweets* harus dikonversi menjadi *dictionary* agar setiap kata unik memiliki ID seperti pada Tabel 3.5. Lalu dikonversi lagi dalam format *Bag of Words* sebagaimana ditunjukkan pada Tabel 3.6 agar dapat memudahkan LDA dalam mengestimasi distribusi topik.



Gambar 3.7 Diagram Alir Pemodelan Topik

Tabel 3.5 Proses Konversi *Dictionary*

No. Tweets	Tweets	Token	Kata Unik	ID Kata
1	QRIS dapat digunakan di seluruh Indonesia	“QRIS”	“QRIS”	1
		“dapat”	“dapat”	2
		“digunakan”	“digunakan”	3
		“di”	“di”	4
		“seluruh”	“seluruh”	5
		“Indonesia	“Indonesia	6
2	Bank Indonesia merilis QRIS	“Bank”	“Bank”	7
		“Indonesia”	“merilis”	8
		“mengembangkan”		
		“QRIS”		

Tabel 3.6 *Dictionary* ke *Bag of Words*

No. Tweets	ID Kata							
	1	2	3	4	5	6	7	8
1	1	1	1	1	1	1	0	0
2	1	0	0	0	0	1	1	1

Setelah dirubah menjadi format *Bag of Words*, ditentukan rentang jumlah topik, *alpha*, dan *passes* yang hendak digunakan dalam *training* model LDA. Lalu perlu juga ditentukan metrik apa saja yang akan digunakan dalam evaluasi model LDA. Dalam hal ini, metrik yang akan digunakan adalah C_V dan $C_{U_{Mass}}$. *Pseudocode* di bawah menjelaskan bagaimana proses pemodelan topik menggunakan LDA:

Pemodelan Topik LDA
<p>Deklarasi:</p> <pre> num_topics_ = range(3 to 10) alpha_range = [0.01, 0.1, 1] passes_ = [10, 20, 50, 100] model_id = 0 df_results = ['model_id', 'num_topics', 'alpha', 'passes', 'cv_score', 'umass_score', and 'topics'] </pre>
<pre> START FOR each num_topics in num_topics_range FOR each alpha in alpha_range FOR each passes in passes_range SET lda_model as a new LDA model SET cv_score using 'c_v' SET umass_score using 'u_mass' SET top_topics as an empty list FOR each topic_id and topic_words in lda_model's APPEND the tuple (topic_id, topic_words) to top_topics APPEND a new row to df_results with model_id, num_topics, alpha, passes, cv_score, umass_score, and top_topics INCREMENT model_id by 1 END FOR END FOR END FOR OUTPUT df_results END </pre>

Output dari *code* di atas akan menghasilkan *dataframe* berisi kolom *model_id*, *num_topics*, *alpha*, *passes*, *cv_score*, *umass_score*, dan *top_topics*. Kolom *top_topics* berisikan *list* kumpulan kata dari masing-masing topik yang telah di-generate oleh LDA. Selanjutnya, model dengan nilai koherensi tertinggi akan dipilih untuk proses analisis topik.

Metode *text mining* seperti pemodelan topik tidak dapat menggantikan analisis manusia, tetapi hanya sebagai alat bantu dalam meningkatkan analisis manusia (Debortoli et al., 2016). Kumpulan kata dari topik-topik yang dihasilkan oleh LDA akan diinterpretasi dan dianalisis secara manual. Pemodelan topik pada label positif dilakukan agar dapat menganalisis opini publik terkait topik yang mengarah pada kelebihan QRIS, sedangkan topik-topik pada label negatif juga akan berguna dalam menganalisis terkait hal-hal yang dapat menjadi ancaman dan peluang agar QRIS dapat menjadi lebih baik.



UIN SUNAN AMPEL
S U R A B A Y A

BAB IV

HASIL DAN PEMBAHASAN

4.1. *Data Collection*

Penelitian ini menggunakan data yang diambil dari Twitter. Data yang diambil merupakan *tweets* yang memiliki kata kunci “QRIS” dan *tweets* yang tersedia sejak dirilisnya QRIS pada tanggal 17 Agustus 2019 hingga 17 Agustus 2022. Pengumpulan data memanfaatkan *library* sncscrape dan terkumpul sebanyak 244.708 *tweets*. Rincian jumlah *tweets* setiap bulannya dapat dilihat pada Tabel 4.1. Tabel 4.1 Jumlah *Tweets*

Tahun	Bulan	Jumlah <i>Tweets</i>
2019	(Agustus - Desember)	1.381
2020	(Januari - Desember)	11.110
2021	(Januari - Desember)	155.426
2022	(Januari - Agustus)	76.791
Total		244.708

Jumlah *tweets* pada tahun 2019, 2020, dan 2021 memiliki selisih yang sangat jauh. Hal ini dikarenakan saat QRIS dirilis, masih banyak masyarakat yang belum *aware* dan mengadopsi dengan baik kelebihan QRIS. Saat pandemi berlangsung, dengan adanya regulasi pemerintah Indonesia untuk mengurangi penggunaan uang kertas, adopsi QRIS semakin meningkat.

4.2. *Data Cleaning*

Data cleaning dilakukan dengan *drop* seluruh *tweets* dari *user* yang memiliki *tweets* mengandung kata dari daftar kata-kata yang tidak relevan dengan penelitian dan berpotensi sebagai spam. Kata-kata yang tidak relevan dikumpulkan secara manual. Contoh dari kata-kata yang berpotensi berupa spam seperti: *Netflix*, *iflix*, *spotify*, *premium*, dll. Setelah dilakukan *data cleaning*, didapat total *tweets* sebanyak 73.764 *tweets*.

4.3. Data Preprocessing

Data preprocessing dilakukan untuk mempersiapkan data agar siap diolah. *Data preprocessing* terdiri dari beberapa tahapan. Berikut detail tahapan *data preprocessing* meliputi:

4.3.1. Remove Username, Hashtag, URLs

Pada tahapan awal pra-proses data, dilakukan penghapusan *username*, *hashtag*, dan *link* URL menggunakan regex. Tabel 4.2 menampilkan transformasi sebelum dan sesudah dilakukannya tahapan ini.

Tabel 4.2 Hasil *Remove Username, Hashtag, URLs*

Sebelum <i>Remove Username, Hashtag, URLs</i>	Setelah <i>Remove Username, Hashtag, URLs</i>
Di tempatmu ada yang mirip-mirip seperti ini? *Ibu-ibu pedagang di Pasar Apung Lok Baintan mengalungi #QRIS mereka. https://t.co/yF2eMpkn5j	Di tempatmu ada yang mirip-mirip seperti ini? *Ibu-ibu pedagang di Pasar Apung Lok Baintan mengalungi mereka.
@fxmario Saran gue, QRIS yang dipajang yang connect ke Gobiz atau OVO aja. Biasanya dua aplikasi tersebut kan adanya di hp operasional, bukan owner. Jadi bisa langsung dicek saat itu juga.	Saran gue, QRIS yang dipajang yang connect ke Gobiz atau OVO aja. Biasanya dua aplikasi tersebut kan adanya di hp operasional, bukan owner. Jadi bisa langsung dicek saat itu juga.
@_sdction Mesjid di tempat gw, kotak amalnya ada qr code QRIS, tinggal scan aja kalo gk punya uang kecil atau lupa bawa duit waktu jumatatan	Mesjid di tempat gw, kotak amalnya ada qr code QRIS, tinggal scan aja kalo gk punya uang kecil atau lupa bawa duit waktu jumatatan

4.3.2. Remove Duplicates

Menghapus *tweets* duplikat dilakukan setelah melakukan pembersihan URL agar *tweets* duplikat yang mengandung *link* dapat terdeteksi. Pada tahapan ini *tweets* yang tersisa sebanyak 68.222 *tweets*.

4.3.3. Remove Punctuation

Punctuation atau tanda baca dapat mempengaruhi hasil analisis sentimen pada *tweets* karena dapat memberikan makna tambahan. Tahapan ini dilakukan dengan menggunakan *library* Python string untuk menghapus semua tanda baca dalam *tweets* seperti `'!#$%&\'()*+,-./:;<=>?@[\\]^_`{|}~'`. Tabel 4.3 menampilkan transformasi sebelum dan sesudah dilakukannya *remove punctuation*.

Tabel 4.3 Hasil *Remove Punctuation*

Sebelum <i>Remove Punctuation</i>	Setelah <i>Remove Punctuation</i>
Abis Jum'atan mau ngisi kencleng, lha koq rogoh saku ternyata ga bawa dompet? Eh ternyata ada #QRIS nya! Alhamdulillah.	Abis Jumatn mau ngisi kenclengn lha koq rogoh saku ternyata ga bawa dompet Eh ternyata ada QRIS nya Alhamdulillah
Kami2 yg UMKM pejuang bazaar sekarang ga perlu repot lagi sediain uang kembalian. #QRIS bener2 membantu. Mau pke rekening apapun, e-money apapun tinggal <i>scan</i> , beres. Tercatat semua lagi!	Kami2 yg UMKM pejuang bazaar sekarang ga perlu repot lagi sediain uang kembalian QRIS bener2 membantu Mau pke rekening apapun emoney apapun tinggal <i>scan</i> beres Tercatat semua lagi
Baru ae semalem beli bebek madura pinggir jalan ngasih cash ibu yang jual nanya "nggak pake qris aja mas"	Baru ae semalem beli bebek madura pinggir jalan ngasih cash ibu yang jual nanya nggak pake qris aja mas

4.3.4. *Remove Emoji*

Pada tahapan ini, karakter emoji yang terdapat pada *tweets* diidentifikasi. Lalu dilakukan penghapusan karakter tersebut dari teks *tweets* menggunakan *unicode data*. *Unicode* menyediakan daftar karakter emoji yang dapat dihapus dari *tweets* menggunakan metode penghapusan karakter khusus seperti `re.sub()` pada *library* Python.

Tabel 4.4 Hasil *Remove Emoji*

Sebelum <i>Remove Emoji</i>	Setelah <i>Remove Emoji</i>
🤔 Baru inget bbrp masjid jg kotak amal nya jg sudah ada QRIS Sangat amat future 🙌	Baru inget bbrp masjid jg kotak amal nya jg sudah ada QRIS Sangat amat future
Qris adalah gamechanger Semua toko sac itu kedampak banget berkat qris Yang dulu cari uang kembalian susah banget skrng tinggal <i>scan</i> joss uang masuk ke toko 🤔🤔🤔🤔	Qris adalah gamechanger Semua toko sac itu kedampak banget berkat qris Yang dulu cari uang kembalian susah banget skrng tinggal <i>scan</i> joss uang masuk ke toko
Kalo QRis yg universal enggak kena charge Kalo qris yg di keluarin Gojek ama Grab penjual kena charge 07 per transaksi Sy jualan soalnya 😞	Kalo QRis yg universal enggak kena charge Kalo qris yg di keluarin Gojek ama Grab penjual kena charge 07 per transaksi Sy jualan soalnya

4.3.5. *Case Folding*

Seluruh huruf teks pada *tweets* diubah menjadi huruf kecil agar memiliki makna yang setara. Disajikan Tabel 4.5 berupa transformasi sebelum dan sesudah dilakukannya *case folding*.

Tabel 4.5 Hasil *Case Folding*

Sebelum <i>Case Folding</i>	Setelah <i>Case Folding</i>
GILA PARAH SUKA BGT SAMA QRIS MAKASIH DANA	gila parah suka bgt sama qris makasih dana
SUMPAH W CINTA MATI BANGET SAMA QRIS PENYELAMAT MANUSIA CASHLESS SEPERTIKU	sumpah w cinta mati banget sama qris penyelamat manusia <i>cashless</i> sepertiku
NAH KAYA GINI MANTAP Ga perlu yg pake edc dah yang bisa pake QRIS aja udah ngeganti orang <i>cashless</i> kek gue	nah kaya gini mantap ga perlu yg pake edc dah yang bisa pake qris aja udah ngeganti orang <i>cashless</i> kek gue

4.3.6. Normalization

Tweets berbahasa Indonesia cenderung banyak mengandung bahasa tidak baku seperti *slang*, bahasa gaul, hingga bahasa daerah. Pada tahapan ini dilakukan penanganan kata-kata di luar kosakata atau *out of vocabulary* (OOV) menggunakan Colloquial *Lexicon*. Selain dapat menormalisasi bahasa gaul sehari-hari seperti “bokap” menjadi “bapak”, “gokil” menjadi “gila”, “thx” menjadi “terima kasih” juga dapat menangani singkatan dan *elongated words* yang lazim digunakan dalam sehari-hari seperti “ojol” menjadi “ojek *online*” dan “bbrp” menjadi “beberapa”.

Tabel 4.6 Hasil *Normalization*

Sebelum <i>Normalization</i>	Setelah <i>Normalization</i>
kadang ngebayangin kalau <i>bokap</i> masih ada pasti <i>bakalan</i> takjub dengan kemajuan teknologi jaman sekarang kayak qris ecommerce atau layanan <i>ojol</i> dan gue yakin <i>doi</i> bakalan minta dibeliin martabak telur via gofood minimal seminggu sekali	kadang membayangkan kalau <i>ayah</i> ada pasti <i>akan</i> takjub dengan kemajuan teknologi jaman sekarang kayak qris ecommerce atau layanan <i>ojek online</i> dan aku yakin <i>dia</i> akan meminta dibelikan martabak telur via gofood minimal seminggu sekali
buset musisi jalanan udah pake qris qr <i>code</i> hari gini <i>gokil</i>	buset musisi jalanan udah pake qris qr <i>code</i> hari gini <i>gila</i>
<i>thx</i> qris <i>bbrp</i> kali keluar kota gw gak perlu <i>bw</i> tunai <i>byk</i> di dompet krn bayar apapun <i>skrg cm</i> modal hape doang	<i>terima kasih</i> qris <i>beberapa</i> kali keluar kota aku tidak perlu <i>bawa</i> tunai <i>banyak</i> di dompet karena bayar apapun <i>sekarang cuma</i> modal telepon genggam doang

4.3.7. Tokenization

Pada tahap *tokenization* teks pada *tweets* dikonversi menjadi token-token yang lebih kecil, seperti kata-kata atau frasa, sehingga dapat diolah lebih lanjut.

Teks pada *tweets* akan dibagi-bagi menjadi bagian-bagian yang lebih kecil yang disebut dengan token.

Tabel 4.7 Hasil *Tokenization*

Sebelum <i>Tokenization</i>	Setelah <i>Tokenization</i>
kadang membayangkan kalau ayah ada pasti akan takjub dengan kemajuan teknologi jaman sekarang kayak qris ecommerce atau layanan ojek online dan aku yakin dia akan meminta dibelikan martabak telur via gofood minimal seminggu sekali	['kadang', 'membayangkan', 'kalau', 'ayah', 'ada', 'akan', 'takjub', 'dengan', 'kemajuan', 'teknologi', 'jaman', 'sekarang', 'kayak', 'qris', 'ecommerce', 'atau', 'layanan', 'ojek', 'online', 'aku', 'yakin', 'dia', 'akan', 'meminta', 'dibelikan', 'martabak', 'telur', 'via', 'gofood', 'minimal', 'seminggu', 'sekali']
buset musisi jalanan udah pake qris qr code hari gini gila gopay dirumahaja	['buset', 'musisi', 'jalanan', 'pakai', 'qris', 'qr', 'code', 'hari', 'gini', 'gila', 'gopay', 'dirumahaja']
terima kasih qris beberapa kali keluar kota aku tidak perlu bawa tunai banyak dompet bayar apapun sekarang cuma modal telepon genggam doang	['terima', 'kasih', 'qris', 'beberapa', 'kali', 'keluar', 'kota', 'aku', 'tidak', 'perlu', 'bawa', 'tunai', 'banyak', 'dompet', 'bayar', 'apapun', 'sekarang', 'cuma', 'modal', 'telepon', 'genggam', 'doang']

4.3.8. *Remove Stopwords*

Pada tahapan ini, kata-kata yang umumnya tidak memiliki makna yang signifikan dan sering muncul seperti “dan”, “atau”, “yang”, dan sebagainya dihapuskan. Hal ini guna mengurangi dimensi dari data dan meningkatkan akurasi analisis. Penghapusan dilakukan dengan cara membandingkan setiap kata dalam *tweets* dengan daftar *stopwords*, dan jika ditemukan kata yang termasuk dalam daftar, maka kata tersebut akan dihapus dari *tweets*. Dalam proses ini, *library* yang digunakan yaitu Sastrawi, masih terdapat beberapa *stopwords* yang tidak terdeteksi sehingga harus membuat *list stopwords* tambahan.

Tabel 4.8 Hasil *Remove Stopwords*

Sebelum <i>Remove Stopwords</i>	Setelah <i>Remove Stopwords</i>
['transaksi', 'lebih', 'cepat', 'dan', 'aman', 'pakai', 'qris']	['transaksi', 'lebih', 'cepat', 'aman', 'pakai', 'qris']
['buat', 'pebisnis', 'yang', 'menerima', 'qris', 'tidak', 'bisa', 'tau', 'dananya', 'masuk', 'tidak', 'saran', 'gue', 'stop', 'accepting', 'qris', 'dulu', 'deh', 'kamu', 'can', 'be', 'easily', 'compromised']	['buat', 'pebisnis', 'menerima', 'qris', 'bisa', 'tau', 'dananya', 'masuk', 'tidak', 'saran', 'gue', 'stop', 'accepting', 'qris', 'dulu', 'deh', 'kamu', 'can', 'be', 'easily', 'compromised']

['edc', 'pakai', '3g', 'by', 'the', 'way', 'device', 'yang', 'lebih', 'baru', 'make', 'pos', 'lebih', 'simple', 'pakai', 'qris', 'padahal', 'generasi', 'alpa', 'tidak', 'bakal', 'kenal', 'kartu', 'kartu', 'malas', 'menyimpan', 'kartu', 'fisik']	['edc', 'pakai', '3g', 'by', 'the', 'way', 'device', 'lebih', 'baru', 'make', 'pos', 'lebih', 'simple', 'pakai', 'qris', 'padahal', 'generasi', 'alpa', 'bakal', 'kenal', 'kartu', 'kartu', 'malas', 'menyimpan', 'kartu', 'fisik']
--	---

4.3.9. Drop Tweets

Guna meningkatkan akurasi pelabelan sentimen, penelitian ini hanya menggunakan *tweets* yang memiliki ≥ 12 kata, sebagaimana yang dilakukan oleh (Jha and Mahmoud, 2019). *Tweets* yang memenuhi kriteria tersisa sebanyak 21.536 *tweets*.

4.3.10. Stemming

Tahapan ini dilakukan agar memudahkan proses pelabelan. Kata-kata yang disederhanakan dapat dihitung dan diurutkan lebih mudah. Contohnya, kata-kata seperti “membeli”, “memiliki”, dan “memakai” dapat diubah menjadi kata dasar “beli”, “milik”, dan “pakai”. Dengan menggunakan *stemming*, kita dapat mengurangi jumlah kata dalam dataset dan meningkatkan akurasi pelabelan *tweets*.

Tabel 4.9 Hasil *Stemming*

Sebelum <i>Stemming</i>	Setelah <i>Stemming</i>
['penerapan', 'qris', 'diharapkan', 'memudahkan', 'masyarakat', 'menggunakan', 'e-wallet', 'transaksi', 'merchant', 'menyediakan', 'qr', 'code']	['terap', 'qris', 'harap', 'mudah', 'masyarakat', 'guna', 'e-wallet', 'transaksi', 'merchant', 'sedia', 'qr', 'code']
['cinta', 'banget', 'sama', 'qris', 'sih', 'memudahkan', 'sebagai', 'pengguna', 'malas', 'bawa', 'tunai', 'disatu', 'sisi', 'mengedukasi', 'penjual', 'untuk', 'bikin', 'melek', 'sama', 'digital', 'banking']	['cinta', 'banget', 'sama', 'qris', 'sih', 'mudah', 'bagai', 'guna', 'malas', 'bawa', 'tunai', 'satu', 'sisi', 'edukasi', 'jual', 'untuk', 'bikin', 'melek', 'sama', 'digital', 'banking']
['buat', 'pebisnis', 'yang', 'menerima', 'qris', 'tidak', 'bisa', 'tau', 'dananya', 'masuk', 'tidak', 'saran', 'gue', 'stop', 'accepting', 'qris', 'dulu', 'deh', 'kamu', 'can', 'be', 'easily', 'compromised']	['buat', 'bisnis', 'yang', 'terima', 'qris', 'tidak', 'bisa', 'tau', 'dananya', 'masuk', 'tidak', 'saran', 'gue', 'stop', 'accepting', 'qris', 'dulu', 'deh', 'kamu', 'can', 'be', 'easily', 'compromised']

4.4. Analisis Sentimen

4.4.1. Pelabelan Sentimen

Label sentimen menggunakan VADER dengan InSet *Lexicon* didapat dari skor *compound* berdasarkan bobot kata pada InSet *Lexicon*. Terdapat tiga label yang dihasilkan yakni positif, netral, dan negatif. Tabel 4.10 menampilkan cuplikan hasil pelabelan sentimen.

Tabel 4.10 Hasil Pelabelan

<i>Tweets</i>	<i>Skor Compound</i>	<i>Label</i>
Sekarang bersedekah ke SR makin praktis dan mudah dengan QRIS . cukup: - Masuk Aplikasi - <i>Scan</i> - Ketik nominal transfer - Selesai	0.0141	Positif
UMKM yg pake QRIS hati2 ya. Yg jual ikan di pasar tadi kena tipu 500rb karena yg beli tunjukkan bukti transfer/QRIS palsu dan yg jual ga cek dulu.	-0.0750	Negatif
Seingatku: - “QRIS” dibacanya “KRIS” - “Croissant” dibacanya “Kwasong” - “Hyundai” dibacanya “Hyonde”	0.0	Netral

4.4.2. Data Validation

Validasi pelabelan dilakukan untuk mengetahui validitas pelabelan yang dihasilkan oleh VADER. Validasi pelabelan dilakukan secara manual oleh manusia (partisipan yang memenuhi kualifikasi). Agar hasil validasi lebih objektif, satu *tweets* divalidasi oleh tiga partisipan.

Data yang digunakan untuk validasi adalah sampel dari seluruh *tweets* yang telah dilabeli oleh VADER. Teknik sampel yang digunakan adalah *random sampling*. Data sampel *tweets* diambil secara *random* sebanyak 300 dari 21.536 *tweets* yang terdiri atas 14.555 *tweets* dengan label positif, 6.082 *tweets* dengan label negatif, dan 899 *tweets* dengan label netral. Pengambilan sampel sebanyak 300 *tweets* dapat merepresentasi dari total 21.536 *tweets* karena memiliki *confidence level* 95% dan *margin of error* 5.62%. Dengan kata lain, jika kita melakukan pengambilan sampel berulang-ulang, maka 95% dari hasil tersebut akan menghasilkan nilai yang benar atau mendekati nilai yang sebenarnya (Singh and Masuku, 2014).

Tabel 4.11 menampilkan bagaimana proses validasi antara pelabelan manusia dan VADER. Jika terdapat lebih dari satu label dari manusia yang sama dengan label yang dihasilkan oleh VADER, maka pelabelan tersebut dianggap terprediksi dengan benar atau valid.

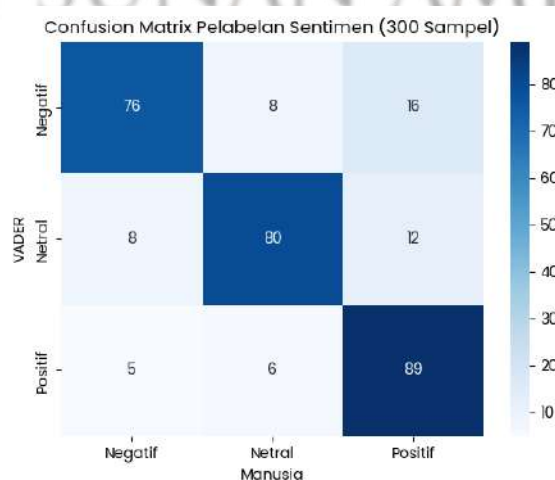
Tabel 4.11 *Data Validation*

<i>Tweets</i>	Pelabelan Manusia			VADER	<i>Valid/Invalid</i>
	1	2	3		
Wow sekarang jajan kue cubit gerobakan harganya goceng bayarnya udah bisa pake QRIS gile I love technology	Positif	Positif	Positif	Positif	Valid
Orang Indonesia jaman dulu dah pake QRIS	Netral	Negatif	Netral	Positif	Invalid
guis aku td abis dr pasar, wkwkwkw lucu bgt skrg kang sayur udh pake qris. jd inget temen ku pas di sekolah jualan makroni keliling kelas, bayarnya bisa <i>cashless</i> jd ga ada alesan buat ga beli (biasanya suka ada yg ngeles ga punya cash) WKWKWKWKWK	Positif	Negatif	Positif	Positif	Valid
Banyak kasir yang terima pembayaran QRIS tapi mereka GA BISA cek dananya masuk atau engga secara REALTIME. It's a BIG mistake! Mereka ini sasaran empuk para penipu sih. Biasanya yang bisa lihat dana masuk itu hanya owner. Para kasir ini diminta foto bukti transfer doang.	Negatif	Netral	Negatif	Negatif	Valid
Kalo aku sekarang jadi kebiasaan pake QRIS, tapi kdang bingung sama rumah makan yg gak cek/foto bukti pembayarannya. Biasanya aku pasti paksa kasirnya buat cek lagi ini mbak tolong, takutnya saya bohong... Kasirnya selalu bilang "percaya sama Mas" Akhirnya ciuman	Netral	Negatif	Negatif	Negatif	Valid
Work! Sebagai anak rantau yang kerja di Jakarta, aku terkagum dengan kantin yang bisa bayar pake QRIS. I mean, TUKANG AYAM PENYET BISA LAYANIN CASHLESS????? KITA HIDUP DI TAHUN 2322 ATAU GIMANA?	Positif	Positif	Positif	Positif	Valid

Baru banget kejadian, beli soto mie dipinggir jalan pake qris. Pas makanan jadi bilang ke amangnya mang ini udah dibayar ya lewat qris sambil nunjukin hp, si amangny cuma iya-iya aja ga liat/cek dulu Langsung shock, takut ada yang jahat ke usaha amangnya gimana	Negatif	Negatif	Positif	Negatif	Valid
Bahkan di Indo, tukang tahu bulat jg udh byk yg sediain QRIS, sungguh futuristik	Positif	Positif	Positif	Positif	Valid
Malem-malem setengah mati cari atm, ternyata ankringannya terima qris	Positif	Netral	Positif	Positif	Valid
Ralat juga, QRIS dari BI, bukan dari depkeu. Penjelasan yang lebih paripurna:	Netral	Netral	Netral	Netral	Valid

4.4.3. Evaluation

Berdasarkan hasil validasi pelabelan yang dilakukan secara manual, didapat *confusion matrix* yang ditampilkan pada Gambar 4.1. VADER cenderung tidak tepat dalam melabeli *tweets* yang sebenarnya positif. Hal ini diduga karena masyarakat cenderung membuat *tweets* terkait QRIS menggunakan kata-kata negatif yang sebetulnya memiliki konteks menyampaikan kelebihan QRIS. Contoh *tweets* tersebut seperti “Eh sama dong, balik ke indo gih, skrg beli gorengan udah bisa pake QRIS loh jadi gausah **ribet** pake tunai dan **ribet** cari kembalian wkwk” dan “Di sekitar kantor gue bnyk warteg dan warung2 **kecil** trs semua pake qris, ktnya **gk** terima cash krn **ribet** kalo pas lg rame.”. Sedangkan VADER menentukan pelabelan sentimen berdasarkan bobot kata-kata yang terdapat pada *tweets* tersebut.



Gambar 4.1 *Confusion Matrix* Hasil Evaluasi

Accuracy, *precision*, *recall*, dan *F1 Score* dapat dilihat pada Tabel 4.12. Pelabelan sentimen menggunakan VADER dengan InSet *lexicon* pada *dataset tweets* QRIS didapat *accuracy* sebesar 81,66% dimana termasuk *accuracy* yang bagus, jika melihat penelitian serupa seperti pelabelan sentimen penanganan COVID-19 sebesar 81% (Ependi et al., 2023), pelebalan sentimen komentar video sebesar 79,78% (D, 2019), dan pelabelan sentimen layanan pesan antar makanan sebesar 71% (Yulivia, 2023).

Terlihat bahwa nilai *precision* lebih tinggi dari *accuracy*, *recall*, dan *F1 Score*. *Precision* mengukur seberapa akurat VADER dalam mengidentifikasi sentimen positif atau negatif pada data, dan nilai *precision* yang lebih tinggi menunjukkan bahwa VADER memiliki performa yang lebih baik dalam mengidentifikasi sentimen yang sebenarnya.

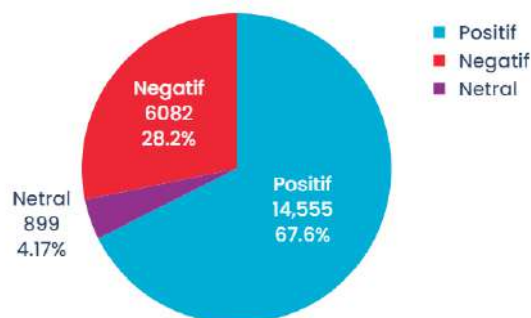
Tabel 4.12 Hasil Evaluasi

<i>Measure</i>	<i>Results</i>
<i>Accuracy</i>	81,66%
<i>Precision</i>	82,32%
<i>Recall</i>	81,66%
<i>F1 Score</i>	81,69%

4.4.4. Analisis Hasil Sentimen

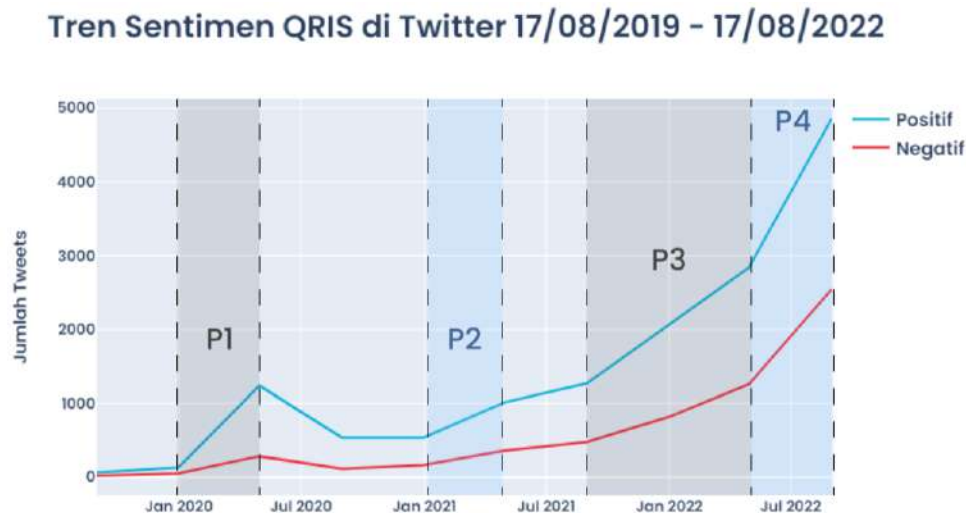
Berdasarkan hasil pelabelan sentimen VADER pada data *tweets*, didapat bahwa masyarakat Indonesia cenderung memiliki sentimen positif. Distribusi sentimen masyarakat Indonesia terhadap QRIS pada Twitter dapat dilihat pada Gambar 4.2.

Distribusi Sentimen Tweets QRIS



Gambar 4.2 Distribusi Sentimen QRIS

Adapun tren QRIS sejak dirilisnya QRIS yakni 17 Agustus 2019 hingga 17 Agustus 2022 terus mengalami peningkatan. Analisis tren ini dilihat dari *tweets* teratas dengan jumlah *likes* paling banyak.



Gambar 4.3 Tren QRIS Selama Tiga Tahun

Pada lonjakan pertama di bulan Desember 2019 hingga April 2020 didominasi oleh *tweets* dari berbagai akun resmi Bank Jatim, Kompas, detik.com yang menginformasikan, mengenalkan, memberi edukasi tentang QRIS kepada masyarakat. Tak hanya itu, pada masa ini juga muncul banyak promo diskon dan *cashback* dari berbagai *merchant* yang merupakan bentuk pengenalan dan upaya meningkatkan adopsi QRIS. Pada masa ini juga resmi diumumkan kasus COVID-19 pertama di Indonesia dan diberlakukannya Pembatasan Sosial Berskala Besar (PSBB) sehingga masa ini merupakan masa awal pengenalan dan adopsi QRIS bagi masyarakat Indonesia.

Lonjakan kedua yaitu sekitar bulan Desember 2020 hingga April 2021, mulai banyak masyarakat Indonesia yang berbagi pengalaman mereka menggunakan QRIS. Didukung dengan fakta bahwa pada masa ini diberlakukan Pemberlakuan Pembatasan Kegiatan Masyarakat (PPKM), masyarakat Indonesia terpaksa untuk mengurangi penggunaan uang kertas sehingga adopsi penggunaan QRIS pun meningkat. Hal ini membuktikan bahwa QRIS sangat berperan penting dalam transaksi kehidupan sehari-hari di era pandemi. Di masa ini juga masih tetap banyak *tweets merchant* yang mengadakan promo untuk pembayaran QRIS.

Lonjakan ketiga pada Agustus 2021 hingga April 2022, *tweets* teratas terkait QRIS didominasi oleh banyak forum diskusi netizen yang menyampaikan

Tabel 4.13 Daftar 10 Kata Teratas Sentimen Positif

No.	Kata	Frekuensi	No.	Kata	Frekuensi
1	bayar	6.837	6	scan	2.203
2	indonesia	2.502	7	guna	1.803
3	bank	2.305	8	bisa	1.713
4	dana	2.256	9	tunai	1.586
5	transaksi	2.220	10	uang	1.555

Berdasarkan *wordcloud* untuk *tweets* berlabel positif, dapat ditemukan beberapa kata yang sering digunakan oleh masyarakat Indonesia. Kata-kata tersebut diurutkan berdasarkan frekuensi tertinggi dan dapat memberikan gambaran mengenai hal-hal apa yang paling disukai atau diapresiasi oleh masyarakat Indonesia selama menggunakan QRIS. Kata pertama yang paling sering digunakan adalah “bayar”, menandakan bahwa sudah banyak masyarakat Indonesia yang melakukan pembayaran menggunakan QRIS. Selanjutnya, kata “Indonesia” juga sering disebutkan, menunjukkan bahwa pengguna QRIS bangga dengan produk teknologi dalam negeri yang canggih dan bermanfaat dalam kehidupan sehari-hari.

Selain itu, kata “bank”, “dana”, “transaksi”, “scan”, “guna”, dan “bisa” juga cukup dominan digunakan. Kata-kata tersebut dapat diinterpretasikan bahwa banyak masyarakat Indonesia yang melakukan transaksi dengan QRIS menggunakan layanan perbankan dan pembayaran digital lain seperti dompet digital, salah satu yang cukup terkenal di Indonesia yaitu Dana. Masyarakat Indonesia merasa QRIS sangat berguna dan mudah digunakan dalam bertransaksi karena praktis hanya dengan *scan* QR. Terakhir, kata “tunai” dan “uang” juga disebutkan, pengguna QRIS sangat mengapresiasi kemudahan pembayaran digital dengan tidak perlu membawa uang tunai yang berisiko dicuri atau hilang, dan juga penipuan uang palsu.

Secara keseluruhan, analisis sentimen QRIS pada Twitter dengan label positif memberikan gambaran mengenai kata-kata yang sering digunakan oleh pengguna QRIS di Indonesia. Kata-kata tersebut dapat merepresentasi bahwa pembayaran dan *scan* kode QR sangat disukai oleh pengguna karena praktis, dan banyak pengguna yang memanfaatkan QRIS untuk melakukan transaksi dan

menggunakan layanan perbankan serta pembayaran digital lainnya. Ketika membandingkan pembayaran *offline* dan *online*, *mobile payment* memang memiliki banyak keunggulan dari ubikuitas, mobilitas, kenyamanan, dan efisiensi (Liu et al., 2015). Adapun prinsip dari QRIS oleh Bank Indonesia yakni “Cemumuah” yang merupakan singkatan dari cepat, murah, mudah, aman, dan handal (Bank Indonesia, 2023). Dengan banyaknya *tweets* dengan sentimen positif, dapat disimpulkan bahwa QRIS semakin populer di Indonesia karena memberikan kemudahan dan semakin banyak digunakan oleh masyarakat dalam kehidupan sehari-hari.

B. Label Negatif

Hasil analisis sentimen QRIS pada Twitter untuk label negatif mengungkap bahwa pengguna sering mengeluhkan berbagai masalah terkait QRIS. Dalam *wordcloud* pada Gambar 4.5 yang diurutkan berdasarkan frekuensi tertinggi, terlihat bahwa kata “bayar”, “uang”, dan “bank” menjadi kata-kata yang paling banyak disebutkan, mengindikasikan bahwa sebagian besar keluhan berupa permasalahan pembayaran dan perbankan.



Gambar 4.5 Wordcloud Sentimen Negatif

Tabel 4.14 Daftar 10 Kata Teratas Sentimen Negatif

No.	Kata	Frekuensi	No.	Kata	Frekuensi
1	bayar	2.831	6	masuk	605
2	uang	875	7	indonesia	601
3	bank	756	8	tidak	569
4	gabisa	623	9	tunai	526
5	beli	616	10	transfer	515

Kata-kata lain yang sering muncul dalam kategori negatif adalah kata “beli”, “masuk”, “tunai”, dan “transfer” juga sering disebutkan oleh pengguna yang mengalami kesulitan saat melakukan transaksi menggunakan QRIS. Selain itu, terdapat kata “gabisa” atau “tidak bisa”, yang dapat dikaitkan dengan adanya masalah teknis atau kegagalan transaksi. Adapun kata “gabisa” paling banyak digunakan diikuti dengan kata-kata sebagai berikut:

Tabel 4.15 Tabel Kata “Tidak Bisa”

Gabungan Kata	Jumlah
Tidak bisa / Gak bisa / Ga bisa / Gabisa pakai	119
Tidak bisa / Gak bisa / Ga bisa / Gabisa QRIS	86
Tidak bisa / Gak bisa / Ga bisa / Gabisa transfer	56
Tidak bisa / Gak bisa / Ga bisa / Gabisa dipakai	37
Tidak bisa / Gak bisa / Ga bisa / Gabisa scan	37

Dari hasil analisis ini, dapat disimpulkan bahwa masih terdapat beberapa masalah teknis dan kendala yang dialami oleh pengguna QRIS. Oleh karena itu, perlu ditingkatkan perbaikan masalah teknis, memberikan kemudahan serta kenyamanan dalam melakukan transaksi, dan terus memperluas akses QRIS di seluruh Indonesia. Dengan demikian, diharapkan pengguna dapat merasa lebih puas dan percaya dalam menggunakan QRIS sebagai gerbang utama pembayaran digital di Indonesia.

4.5. Pemodelan Topik

Pemodelan topik dilakukan pada masing-masing *tweets* label positif dan negatif. Proses *running* model LDA dilakukan secara iteratif dengan rentang jumlah topik 3-10, rentang *alpha* 0,01, 0,1, dan 1, serta *passes* 10, 20, 50, 100. Hal ini dilakukan guna menemukan model dengan C_V dan $C_{U\ Mass}$ tertinggi.

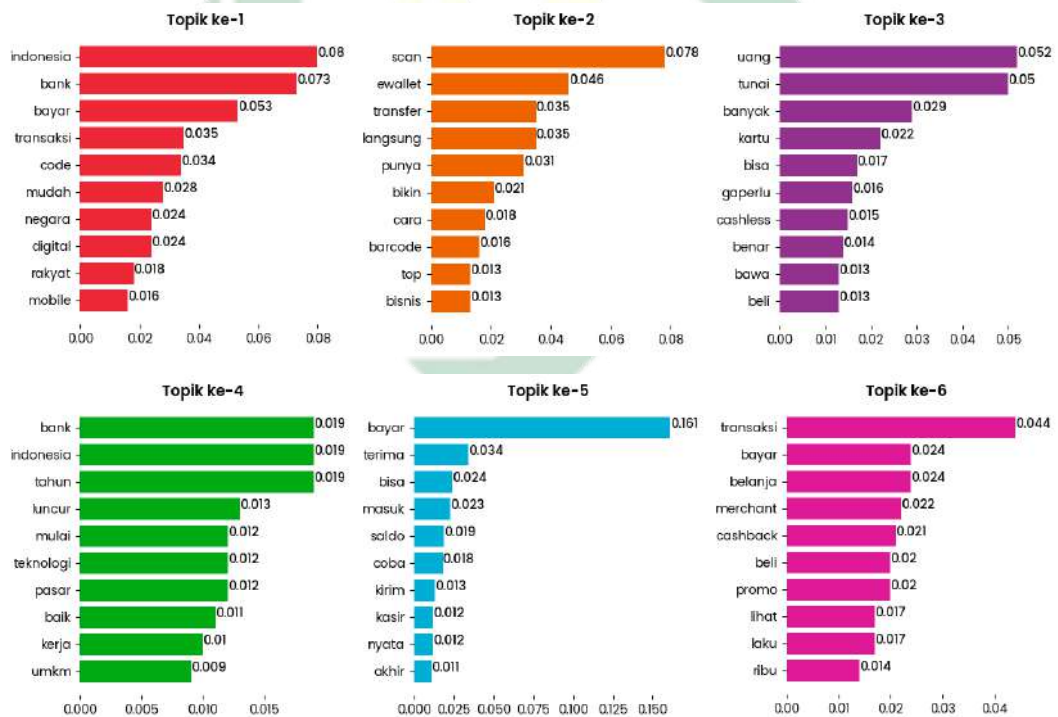
A. Label Positif

Tabel 4.16 menampilkan lima model dengan C_V dan $C_{U\ Mass}$ tertinggi untuk *tweets* dengan label positif. Didapat model dengan enam topik memiliki C_V tertinggi dengan $C_{U\ Mass}$ yang juga relatif tinggi. Penelitian ini mengutamakan seberapa mudah topik untuk diinterpretasikan, sehingga model dengan C_V tertinggi dipilih untuk dianalisis.

Tabel 4.16 Daftar Koherensi Pemodelan Topik *Tweets* Positif

Model ID	Jumlah Topik	α	Passes	C_V	$C_{U\ Mass}$
48	6	1	100	0.488037	-3.460866
67	8	0,1	50	0.479753	-4.333161
71	8	1	50	0.478608	-3.675073
15	4	0,01	50	0.469535	-3.017069
22	4	1	20	0.466051	-3.236001

Pemodelan topik QRIS dengan label positif memberikan hasil yang menarik. Terdapat enam topik yang masing-masing memberikan perspektif berbeda dalam penggunaan QRIS oleh masyarakat Indonesia. Gambar 4.6 menampilkan visualisasi hasil pemodelan topik dilengkapi dengan kata dan bobot kata pada masing-masing topik.



Gambar 4.6 Hasil Pemodelan Topik *Tweets* Positif

Topik pertama mencakup kata-kata seperti “bank”, “bayar”, “transaksi”, “code”, “digital”, dan “mobile” yang terkait dengan pembayaran digital di Indonesia, membuktikan bahwa QRIS dianggap sebagai solusi digital untuk transaksi keuangan yang efisien dan dapat diakses melalui perangkat *mobile*. Transaksi pembayaran di sini tidak hanya merujuk pada transaksi pembayaran jual

beli saja namun juga pada pembayaran zakat, infak, dan sedekah (Prayitno and Fadly, 2022). Kata-kata seperti “rakyat”, “negara”, “Indonesia”, dan “mudah” mengungkap pentingnya kemudahan dan aksesibilitas dalam pembayaran digital bagi seluruh rakyat Indonesia. Topik ini dapat merepresentasi bahwa QRIS dapat memberikan alternatif digital yang efisien untuk membantu orang Indonesia melakukan transaksi secara mudah dan efektif. Salah satu *tweets* yang mewakili topik pertama ini yaitu “*Mantap deh pada era digital sekarang, apa-apa udah pada serba digital, mau bayar ini itu jg pake cara digital. Tinggal tap-tap atau scan QRIS beres deh*”.

Topik kedua mencakup kata-kata seperti “*scan*”, “*e-wallet*”, dan “*transfer*”, yang membahas penggunaan QR code atau QRIS itu sendiri, menggunakan *e-wallet* pada sektor bisnis di Indonesia. Kata kunci seperti “*punya*” dan “*bikin*” mengindikasikan bahwa pengguna QRIS cenderung membuat akun *e-wallet* mereka sendiri untuk menyimpan uang mereka dalam bentuk digital dan melakukan transaksi menggunakan QRIS. Selain itu, kata kunci seperti “*bisnis*” dan “*top*” dapat dikaitkan dengan fakta bahwa QRIS semakin populer di kalangan bisnis dan membantu pengusaha meningkatkan efisiensi dalam melakukan pembayaran atau transaksi dengan pelanggan. Topik ini membuktikan bahwa QRIS dapat menjadi solusi pembayaran yang efisien dan populer bagi bisnis di Indonesia. Berikut salah satu *tweets* seorang pebisnis berbagi pengalamannya selama menggunakan QRIS “*Ini benar. QRIS memudahkan transaksi bisnis kami. Pembeli bisa langsung transfer ke rekening tanpa ribet memasukkan nomor rekening, dari sisi penjual juga nyaman karena langsung ada pemberitahuan adanya dana masuk sekian ke rekening... :)*”.

Topik ketiga mencakup kata-kata seperti “*uang*”, “*tunai*”, dan “*kartu*” yaitu terkait keamanan transaksi dalam transformasi *cashless* di Indonesia. Menunjukkan kebutuhan untuk mengurangi ketergantungan pada uang tunai dalam transaksi. Kata-kata seperti “*bawa*”, “*gaperlu*”, dan “*cashless*” menggambarkan kepraktisan QRIS sebagai alternatif pembayaran digital yang lebih mudah dan aman dibandingkan dengan membawa uang tunai. Topik ini mengungkap bahwa QRIS dapat membantu mengurangi penggunaan uang tunai di Indonesia dan meningkatkan keamanan transaksi. Berikut salah satu *tweets* terkait topik ini “*Skrng*

udah hampir ga pernah bayar ini itu pake cash, pake debit card pun udah jarang banget. Qris the best sih emang, mempermudah hidup makhluk cashless macem hamba dan lebih aman juga”.

Topik keempat mencakup kata-kata seperti “bank”, “Indonesia”, “teknologi”, dan “umkm”, dapat dikaitkan dengan peran Bank Indonesia dalam meningkatkan kemampuan Usaha Mikro Kecil dan Menengah (UMKM) untuk bertransaksi dan memperluas pasar melalui teknologi QRIS. Kata-kata seperti “pasar” dan “baik” mewakili dampak positif penggunaan QRIS bagi UMKM dalam meningkatkan aksesibilitas produk dan layanan bagi pasar mereka. Topik ini merepresentasikan bahwa QRIS dapat memberikan dukungan teknologi yang penting bagi UMKM di Indonesia. Pada penelitian di Kota Pematangsiantar juga telah membuktikan bahwa implementasi QRIS memiliki pengaruh positif terhadap kemajuan UMKM (Hutagalung et al., 2021).

Topik kelima mencakup kata-kata seperti “bayar”, “terima”, dan “saldo” yakni tentang pengalaman positif pengguna QRIS yang dapat membayar dan menerima saldo dengan cepat dan mudah. Pengguna dapat dengan mudah memeriksa saldo mereka. Kata-kata seperti “coba”, “bisa” dan “nyata” berarti QRIS telah terbukti berhasil dalam meningkatkan efisiensi transaksi. Topik ini mengungkap bagaimana QRIS dapat membantu pengguna melakukan transaksi dengan cepat dan mudah, serta memberikan pengalaman yang positif bagi pengguna. Salah satu *tweets* dengan topik ini yaitu “*Sekarang beramal semakin mudah ya, bisa diam-diam dengan memindai QRIS pakai GoPay, OVO, Link aja, langsung terdeteksi nama penerima. Lebih cepat dan praktis”.*

Topik terakhir mencakup kata-kata seperti “transaksi”, “bayar,” “cashback”, “promo”, “belanja”, “beli”, dan “lihat” dapat dikaitkan dengan *benefit* atau keuntungan bagi pengguna dalam menggunakan QRIS berupa *cashback* dan promo khusus dalam melakukan pembayaran pada saat berbelanja. Hal ini membuktikan bahwa QRIS dapat memberikan insentif tambahan bagi pengguna untuk menggunakan pembayaran digital sehingga dapat meningkatkan penggunaan QRIS di Indonesia. Selain itu, kata “*merchant*” mewakili bagaimana para pelaku bisnis dapat menggunakan QRIS untuk menawarkan promo dan *cashback* bagi pelanggan mereka sebagai strategi pemasaran yang efektif. Topik ini menunjukkan bahwa

QRIS dapat menjadi alat pemasaran yang efektif bagi bisnis di Indonesia dan memberikan insentif bagi pengguna untuk menggunakan QRIS. Berikut salah satu *tweets* yang mewakili topik ke-enam “*saya ke Giant Blok M Plaza hari Minggu malam (21/3) bisa bayar pake QRIS. Besoknya dapat cashback 50%* 🤗”.

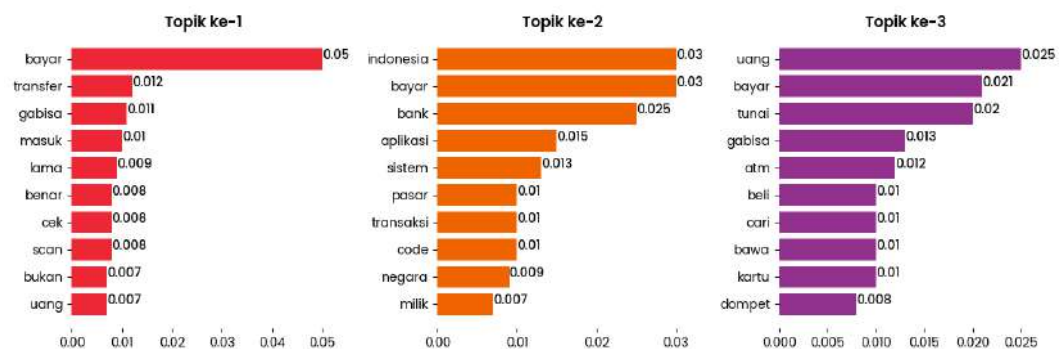
B. Label Negatif

Dalam Tabel 4.17, ditunjukkan lima model teratas untuk *tweets* dengan label negatif. Terdapat satu model dengan tiga topik yang memiliki nilai C_V tertinggi.

Tabel 4.17 Daftar Koherensi Pemodelan Topik *Tweets* Negatif

Model ID	Jumlah Topik	Alpha	Passes	C_V	$C_{U\ Mass}$
7	3	0,1	50	0.383938	-3.837208
12	3	1	100	0.373610	-3.324670
11	3	1	50	0.366320	-3.585475
88	10	0,01	100	0.364659	-5.603811
39	6	0,01	50	0.363716	-5.070396

Dalam pemodelan topik untuk *tweets* dengan label negatif, tiga topik utama berhasil dihasilkan. Visualisasi hasil pemodelan topik beserta kata dan bobot kata pada setiap topik dapat dilihat pada Gambar 4.7. Terdapat beberapa kata kunci yang muncul berkali-kali yang menunjukkan kekhawatiran dan masalah yang dialami oleh masyarakat Indonesia dalam menggunakan QRIS di kehidupan sehari-hari.



Gambar 4.7 Hasil Pemodelan Topik *Tweets* Negatif

Topik pertama yang muncul adalah masalah pembayaran dan transfer menggunakan QRIS. Banyak pengguna yang mengeluhkan kesulitan saat melakukan pembayaran atau transfer menggunakan QRIS. Kata “gabisa”, “bayar”, “transfer”, “scan”, “masuk”, dan “lama” mewakili beberapa masalah yang sering

dihadapi adalah pembayaran atau transfer yang tidak bisa dilakukan, pembayaran atau transfer yang memakan waktu lama, serta kesulitan saat melakukan *scan* QRIS dan cek apakah uang sudah benar-benar masuk ke saldo akun *merchant*. Selain itu, ada juga pengguna yang merasa uang yang mereka bayarkan tidak masuk ke saldo akun *merchant*. Berikut contoh *tweets* keluhan salah satu pelanggan yang menggunakan QRIS kepada salah satu PJSP “*Min saya scan qris dana bisnis.. uang saya sudah terdebit tapi ngak masuk mercant toko dana.. gimana nih min.. bayar belanjaan tapi uangnya ngak masuk ke akun DANA yg punya toko.. solusinya gimana*”.

Topik kedua mencakup beberapa kata kunci seperti “bayar”, “Indonesia”, “bank”, “aplikasi”, “sistem”, “pasar”, “code”, “transaksi”, “negara”, dan “milik”. Banyak *tweets* yang mengeluhkan tentang masalah-masalah aksesibilitas QRIS, seperti aplikasi PJSP yang tidak responsif, akses QRIS di pasar yang kurang memadai, serta masalah teknis lainnya yang menyebabkan kesulitan dalam melakukan transaksi. Salah satu *tweets* yang mewakili topik ini yaitu “*jangan pake qris dari bank juga kalo sinyal lagi jelek. setiap transaksi gagal ke penerima uang, biasanya transaksi berhasil kalo dicek dari mutasi rekening si yang bayar. saldo udah kepotong tapi transaksi ga berhasil*”.

Topik ketiga mencakup beberapa kata kunci seperti “uang”, “bayar”, “tunai”, “gabisa”, “ATM”, “beli”, “cari”, “bawa”, “kartu”, dan “dompet”. Banyak *tweets* yang mengeluhkan tentang sulitnya melakukan pembayaran karena ketersediaan QRIS yang belum merata, masyarakat Indonesia cenderung kesulitan mencari *merchant* yang menerima pembayaran dengan QRIS, sehingga masih terdapat ketergantungan terhadap uang tunai dan tetap harus sedia kartu, dompet, atau mencari ATM untuk melakukan transaksi. Berikut contoh *tweets* yang mewakili topik ini “*Gara2 ATM pada trouble, terpaksa harus jauh2 ke kampus buat ngambil duit aja. Bener sih tinggal di Padang ini masih blm bisa buat belanja cashless. Di Grapari aja kmren harus bayar cash dan gak terima bayar pakai kartu, qris, ataupun transfer*”.

Terkait masalah pembayaran dan transfer pada topik pertama dan masalah aksesibilitas QRIS pada topik kedua, hal ini tidak sepenuhnya bergantung pada sistem QRIS saja namun bisa juga datang dari sistem PJSP yang digunakan atau

koneksi internet kedua pihak baik dari pembeli maupun *merchant* (Natalina et al., 2021). Jika dibandingkan dengan kartu pembayaran, memang merupakan keterbatasan QRIS yang mengharuskan untuk terkoneksi dengan internet agar dapat melakukan transaksi. Tidak dipungkiri fakta bahwa infrastruktur di Indonesia dalam hal koneksi internet masih belum merata pada daerah-daerah tertentu menjadi salah satu penghambat adopsi QRIS dan kendala transaksi QRIS (Tobing et al., 2021). Namun, salah satu latar belakang dikembangkannya QRIS jika berdasarkan data Bank Indonesia di akhir tahun 2021, perbandingan antara jumlah uang elektronik beredar lebih dari dua kali lipat dari debit beredar yaitu 575 juta dan 221 juta (Bank Indonesia, 2021). Diharapkan Bank Indonesia dapat mengembangkan metode pembayaran digital yang dapat dilakukan walaupun sedang tidak terhubung dengan internet, seperti teknologi *Near Field Communication* (NFC) (Igboanusi et al., 2021).

Bank Indonesia meluncurkan QRIS dengan dua jenis, statis dan dinamis. QRIS statis biasanya berupa gambar, stiker, atau poster berisikan QR Code QRIS yang *unique* untuk setiap *merchant* dan bersifat tetap. Cara kerja QRIS statis ini adalah dengan mentranslasi huruf dan angka menjadi bentuk titik (*dots*). QRIS statis tidak bisa diubah sama sekali karena setiap bentuk dan posisi *dot*, merepresentasikan huruf maupun angka tertentu. Berbeda dari QRIS statis, *link* yang tertaut pada QRIS dinamis dapat berubah. Contohnya untuk pembayaran *cashless* pada sebuah toko, QRIS dinamis yang digunakan akan berbeda setiap kali melakukan transaksi sehingga lebih aman. Kelebihan lain dari QRIS dinamis yaitu dapat melacak data. Pelacakan data yang dimaksud di sini antara lain frekuensi QRIS di-*scan*, kapan kode digunakan, lokasi kode digunakan, dan jenis perangkat apa yang digunakan untuk *scan* kode QRIS tertentu (Yovita, 2021).

QRIS yang paling murah dari segi operasional dan paling umum digunakan oleh para UMKM adalah QRIS jenis statis. Tak heran jika sudah terdapat beberapa celah keamanan atau modus penipuan yang merugikan *merchant* atau UMKM. Terdapat beberapa *tweets* viral yang membahas berbagai modus penipuan QRIS yang dialami *merchant* maupun pengguna seperti berikut “UMKM yg pake QRIS hati2 ya. Yg jual ikan di pasar tadi kena tipu 500rb karena yg beli tunjukin bukti transfer/QRIS palsu dan yg jual ga cek dulu”, “Modus “kelebihan 0 waktu transfer

bank dgn screen capture slip transfer palsu” dimodifikasi ke QRIS. Penipu memang laknat”, dan “oh ini stiker QRIS nya ditempel dengan milik penipu ya 🤪 memang perlu edukasi customer kalau nama merchant harus di double check saat pembayaran (misal transaksi merchant A, saat scan QR kok bayarnya jadi ke X)”. Meskipun beberapa *tweets* yang menjelaskan modus penipuan QRIS pada tahun 2022, hingga saat ini (12/04/2023) masih terjadi modus penipuan yang sama yaitu penempelan QRIS palsu, namun untuk kasus ini berkedok sedekah pada berbagai lokasi masjid (Noviansah, n.d.). Hal ini merupakan tugas Bank Indonesia selaku pengembang dan jajaran PJSP terdaftar untuk memperbaiki QRIS statis dalam sisi keamanan dan lebih sigap dalam menanggapi jenis penipuan baru sehingga tidak harus menunggu viral terlebih dahulu agar dapat ditindaklanjuti. Di sisi lain Bank Indonesia juga tetap harus memperluas edukasi keamanan transaksi pada pengguna yang memiliki literasi keuangan dan literasi digital rendah.

Solusi yang saat ini dapat dilakukan untuk para UMKM dalam meminimalisir modus penipuan dan menjaga keberhasilan transaksi adalah dengan selalu siap terhubung dengan koneksi internet dan selalu memeriksa pada setiap transaksi, apakah saldo sudah masuk pada akun atau tidak. Akan lebih baik jika UMKM yang memiliki biaya lebih untuk *upgrade* ke QRIS dinamis sehingga pembayaran menjadi lebih aman. Sebagai pengguna QRIS di sisi pembayar juga harus cermat dalam melakukan transaksi mewaspadaai kasus penipuan yang disebabkan oleh oknum yang menempelkan stiker QRIS palsu (bukan QRIS asli tujuan *merchant*).

Terkait minimnya ketersediaan QRIS pada *merchant* atau pasar tradisional pada topik ketiga, karena dataset *tweets* yang digunakan merupakan *tweets* yang tersedia dalam tiga tahun terakhir, dimana juga beriringan dengan upaya pemerintah untuk melakukan transformasi digital di sektor pasar tradisional, UMKM dan *merchant* seluruh Indonesia. Faktanya, sampai saat ini proses digitalisasi masih terus berlanjut. Kabarnya, dilansir dari idxchannel.com (24/03/2023), seluruh pasar tradisional di Depok sudah dapat menerima pembayaran melalui QRIS (Sandi, 2023)

BAB V PENUTUP

5.1. Kesimpulan

Setelah melakukan berbagai tahapan penelitian, dapat disimpulkan hasil-hasil yang telah ditemukan dari penelitian ini yaitu:

1. Hasil analisis sentimen QRIS menggunakan VADER membuktikan bahwa sentimen masyarakat Indonesia terhadap QRIS cenderung positif dengan 67,6% sentimen positif, 28,2% sentimen negatif, dan 4,17% sentimen netral. Pelabelan sentimen menggunakan VADER dan InSet memiliki performa yang baik dengan akurasi 81,66%. Namun, cenderung tidak tepat dalam melabeli *tweets* yang sebenarnya positif karena masyarakat sering kali membuat *tweets* terkait QRIS dengan menggunakan kata-kata negatif yang sebetulnya memiliki konteks menyampaikan kelebihan QRIS.
2. Hasil analisis pemodelan topik QRIS menggunakan LDA berdasarkan pelabelan sentimen menggunakan VADER menunjukkan bahwa pada *tweets* positif, LDA berhasil menghasilkan enam topik dengan nilai C_V 0.488037 dan $C_{U\ Mass}$ -3.460866 yaitu tentang pembayaran digital, penggunaan QRIS dengan *e-wallet* pada sektor bisnis, keamanan transaksi, implementasi QRIS pada UMKM, pengalaman pengguna, serta *benefit* penggunaan QRIS. Sedangkan pada *tweets* negatif, LDA berhasil menghasilkan tiga topik dengan nilai C_V 0.383938 dan $C_{U\ Mass}$ -3.837208 yaitu tentang masalah pembayaran dan transfer, aksesibilitas QRIS, dan ketersediaan QRIS.

5.2. Saran

Dalam penelitian ini, tentu tidak luput dari kekurangan-kekurangan yang perlu diperbaiki dan batasan permasalahan yang dapat dieksplorasi. Berikut rekomendasi dan saran yang dapat dilakukan pada penelitian berikutnya:

1. Penggunaan VADER dengan InSet *lexicon* tidak dapat memahami konteks *tweets* QRIS yang mengandung banyak kata negatif namun sebetulnya memiliki konteks positif. Hal ini dapat menjadi celah untuk penelitian selanjutnya.

2. Melakukan analisis sentimen tidak hanya dengan label positif, negatif, dan netral saja tetapi juga emosi pengguna seperti yang telah dilakukan (Mathayomchan et al., 2022; Mohamed Ridhwan and Hargreaves, 2021) dalam menganalisis sentimen terkait COVID-19.
3. Menggunakan dataset dari *platform* lain untuk mengetahui opini publik terkait QRIS.
4. Menambah metrik koherensi lain dalam menentukan pemodelan topik terbaik.



DAFTAR PUSTAKA

- Agustina, A., 2017. Analisis Dan Visualisasi Suara Pelanggan Pada Pusat Layanan Pelanggan Dengan Pemodelan Topik Menggunakan Latent Dirichlet Allocation (LDA) Studi Kasus: PT Petrokimia Gresik.
- Aho, A.V., 1990. CHAPTER 5 - Algorithms for Finding Patterns in Strings, in: Van leeuwen, J. (Ed.), Algorithms and Complexity, Handbook of Theoretical Computer Science. Elsevier, Amsterdam, pp. 255–300. <https://doi.org/10.1016/B978-0-444-88071-0.50010-2>
- Albrecht, J., Ramachandran, S., Winkler, C., 2020. Blueprints for Text Analytics Using Python. O'Reilly Media, Inc.
- Aletras, N., Stevenson, M., 2013. Evaluating Topic Coherence Using Distributional Semantics. Presented at the International Conference on Computational Semantics.
- Ali, T., Omar, B., Soulimane, K., 2022. Analyzing tourism reviews using an LDA topic-based sentiment analysis approach. *Methods X* 9. <https://doi.org/10.1016/j.mex.2022.101894>
- Aliyah Salsabila, N., Ardhito Winatmoko, Y., Akbar Septiandri, A., Jamal, A., 2018. Colloquial Indonesian Lexicon, in: 2018 International Conference on Asian Language Processing (IALP). Presented at the 2018 International Conference on Asian Language Processing (IALP), pp. 226–229. <https://doi.org/10.1109/IALP.2018.8629151>
- Anastasia, S., Budi, I., 2016. Twitter sentiment analysis of online transportation service providers. Presented at the 2016 International Conference on Advanced Computer Science and Information Systems (ICACSIS), IEEE, pp. 359–365.
- Attabi, A.W., Muflikhah, L., Fauzi, M.A., 2018. Penerapan Analisis Sentimen untuk Menilai Suatu Produk pada Twitter Berbahasa Indonesia dengan Metode Naïve Bayes Classifier dan Information Gain. *J Pengemb Teknol Inf Dan Ilmu Komput* 2, 4548–4554.
- Aveni, T., Roest, J., 2017. China's Alipay and WeChat Pay: Reaching Rural Users (Brief). World Bank, Washington, DC. <https://doi.org/10.1596/30112>
- Balch, O., 2020. Digital payments: how COVID-19 sped up adoption [WWW Document]. Raconteur. URL <https://www.raconteur.net/finance/payments/digital-payments-covid-19/> (accessed 1.14.23).
- Bank Indonesia, 2023. QR Code Indonesian Standard (QRIS) [WWW Document]. URL <https://www.bi.go.id/QRIS/default.aspx> (accessed 2.2.23).
- Bank Indonesia, 2021. Statistik Sistem Pembayaran dan Infrastruktur Pasar Keuangan (SPIP) [WWW Document]. URL <https://www.bi.go.id/id/statistik/ekonomi-keuangan/spip/Default.aspx> (accessed 3.31.23).
- Barde, B.V., Bainwad, A.M., 2017. An overview of topic modeling methods and tools, in: 2017 International Conference on Intelligent Computing and Control Systems (ICICCS). Presented at the 2017 International Conference on Intelligent Computing and Control Systems (ICICCS), pp. 745–750. <https://doi.org/10.1109/ICCONS.2017.8250563>

- Bird, S., Klein, E., Loper, E., 2009. *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit*. O'Reilly Media, Inc.
- Blei, D.M., Ng, A.Y., Jordan, M.I., 2003. Latent dirichlet allocation. *J. Mach. Learn. Res.* 3, 993–1022.
- Bradley, A.P., 1997. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognit.* 30, 1145–1159. [https://doi.org/10.1016/S0031-3203\(96\)00142-2](https://doi.org/10.1016/S0031-3203(96)00142-2)
- D, C.V., 2019. Hybrid approach: naive bayes and sentiment VADER for analyzing sentiment of mobile unboxing video comments. *Int. J. Electr. Comput. Eng. IJECE* 9, 4452–4459. <https://doi.org/10.11591/ijece.v9i5.pp4452-4459>
- Das, C., Sahoo, A.K., Pradhan, C., 2022. Chapter 12 - Multicriteria recommender system using different approaches, in: Mishra, S., Tripathy, H.K., Mallick, P.K., Sangaiah, A.K., Chae, G.-S. (Eds.), *Cognitive Big Data Intelligence with a Metaheuristic Approach, Cognitive Data Science in Sustainable Computing*. Academic Press, pp. 259–277. <https://doi.org/10.1016/B978-0-323-85117-6.00011-X>
- Das, S., Cakmak, U.M., 2018. *Hands-On Automated Machine Learning: A beginner's guide to building automated machine learning systems using AutoML and Python*. Packt Publishing Ltd.
- Destarani, A.R., Slamet, I., Subanti, S., 2019. Trend Topic Analysis using Latent Dirichlet Allocation (LDA) (Study Case: Denpasar People's Complaints Online Website). *J. Ilm. Tek. Elektro Komput. Dan Inform.* 5, 50–58. <https://doi.org/10.26555/jiteki.v5i1.13088>
- Düntsch, I., Gediga, G., 2020. Indices for rough set approximation and the application to confusion matrices. *Int. J. Approx. Reason.* 118, 155–172. <https://doi.org/10.1016/j.ijar.2019.12.008>
- Ekaputri, A.P., Akbar, S., 2022. Financial News Sentiment Analysis using Modified VADER for Stock Price Prediction, in: *2022 9th International Conference on Advanced Informatics: Concepts, Theory and Applications (ICAICTA)*. Presented at the 2022 9th International Conference on Advanced Informatics: Concepts, Theory and Applications (ICAICTA), pp. 1–6. <https://doi.org/10.1109/ICAICTA56449.2022.9932925>
- Eksa Permana, M., Ramadhan, H., Budi, I., Budi Santoso, A., Kresna Putra, P., 2020. Sentiment Analysis and Topic Detection of Mobile Banking Application Review, in: *2020 Fifth International Conference on Informatics and Computing (ICIC)*. Presented at the 2020 Fifth International Conference on Informatics and Computing (ICIC), pp. 1–6. <https://doi.org/10.1109/ICIC50835.2020.9288616>
- Ependi, U., Aliya, S., Wibowo, A., 2023. Sentiment Analysis of Covid-19 Handling in Indonesia Based on Lexicon Weighting. *J. Sisfokom Sist. Inf. Dan Komput.* 12, 76–82. <https://doi.org/10.32736/sisfokom.v12i1.1615>
- Etaiwi, W., Naymat, G., 2017. The Impact of applying Different Preprocessing Steps on Review Spam Detection. *Procedia Comput. Sci.* 113, 273–279. <https://doi.org/10.1016/j.procs.2017.08.368>
- Fadhillah, I.F.N., Herdiani, A., Astuti, W., 2019. Analisis Sentimen Berbasis Leksikon InSet Terhadap Partai Politik Peserta Pemilu 2019 Pada Media Sosial Twitter. *EProceedings Eng.* 6.

- Fauziah, Y., Yuwono, B., Aribowo, A.S., 2021. Lexicon Based Sentiment Analysis in Indonesia Languages : A Systematic Literature Review. *RSF Conf. Ser. Eng. Technol.* 1, 363–367. <https://doi.org/10.31098/cset.v1i1.397>
- Febrianta, M.Y., Widiyanesti, S., Ramadhan, S.R., 2021. Analisis Ulasan Indie Video Game Lokal pada Steam Menggunakan Analisis Sentimen dan Pemodelan Topik Berbasis Latent Dirichlet Allocation. *J. Animat. Games Stud.* 7, 117–144. <https://doi.org/10.24821/jags.v7i2.5162>
- Firdaus, R., Asror, I., Herdiani, A., 2021. Lexicon-based sentiment analysis of Indonesian language student feedback evaluation. *Indones. J. Comput. Indo-JC* 6, 1–12.
- Hakim, A., 2018. Klasifikasi Sentimen Terhadap Bukalapak dengan Menggunakan Metode Naïve Bayes Classifier (skripsi). Universitas Islam Negeri Sultan Syarif Kasim Riau. https://doi.org/10.10.20BAB%20V_2018464TIF.pdf
- Han, J., Kamber, M., Pei, J., 2012. 13 - Data Mining Trends and Research Frontiers, in: Han, J., Kamber, M., Pei, J. (Eds.), *Data Mining (Third Edition)*, The Morgan Kaufmann Series in Data Management Systems. Morgan Kaufmann, Boston, pp. 585–631. <https://doi.org/10.1016/B978-0-12-381479-1.00013-7>
- Hartini, Y.S., Lefanska, A.B.P., Ursia, A.A., Prasetyo, D.A.B., Sugiharto, B., 2022. *Prosiding Seminar Nasional Sanata Dharma Berbagi “Pengembangan, Penerapan Dan Pendidikan ‘Sains Dan Teknologi’ Pasca Pandemi.”* Sanata Dharma University Press.
- Hunter, J.D., 2007. Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* 9, 90–95. <https://doi.org/10.1109/MCSE.2007.55>
- Hutagalung, R.A., Nainggolan, P., Panjaitan, P.D., 2021. Analisis Perbandingan Keberhasilan UMKM Sebelum Dan Saat Menggunakan Quick Response Indonesia Standard (QRIS) Di Kota Pematangsiantar. *J. Ekuilnomi* 3, 94–103. <https://doi.org/10.36985/ekuilnomi.v3i2.260>
- Hutto, C., Gilbert, E., 2014. VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. *Proc. Int. AAAI Conf. Web Soc. Media* 8, 216–225. <https://doi.org/10.1609/icwsm.v8i1.14550>
- Igboanusi, I.S., Dirgantoro, K.P., Lee, J.-M., Kim, D.-S., 2021. Blockchain side implementation of Pure Wallet (PW): An offline transaction architecture. *ICT Express* 7, 327–334. <https://doi.org/10.1016/j.icte.2021.08.004>
- Jha, N., Mahmoud, A., 2019. Mining non-functional requirements from App store reviews. *Empir. Softw. Eng.* 24, 3659–3695. <https://doi.org/10.1007/s10664-019-09716-7>
- Keerthi Kumar, H.M., Harish, B.S., 2018. Classification of Short Text Using Various Preprocessing Techniques: An Empirical Evaluation: *Proceedings of the 5th ICACNI 2017, Volume 3*, in: *Advances in Intelligent Systems and Computing*. pp. 19–30. https://doi.org/10.1007/978-981-10-8633-5_3
- Koto, F., Rahmaningtyas, G.Y., 2017. Inset lexicon: Evaluation of a word list for Indonesian sentiment analysis in microblogs, in: *2017 International Conference on Asian Language Processing (IALP)*. Presented at the 2017 International Conference on Asian Language Processing (IALP), pp. 391–394. <https://doi.org/10.1109/IALP.2017.8300625>

- Lappeman, J., Clark, R., Evans, J., Sierra-Rubia, L., Gordon, P., 2020. Studying social media sentiment using human validated analysis. *MethodsX* 7, 100867. <https://doi.org/10.1016/j.mex.2020.100867>
- Liu, B., 2012. *Sentiment Analysis and Opinion Mining, Synthesis Lectures on Human Language Technologies*. Springer International Publishing, Cham. <https://doi.org/10.1007/978-3-031-02145-9>
- Liu, J., Kauffman, R.J., Ma, D., 2015. Competition, cooperation, and regulation: Understanding the evolution of the mobile payments technology ecosystem. *Electron. Commer. Res. Appl., Contemporary Research on Payments and Cards in the Global Fintech Revolution* 14, 372–391. <https://doi.org/10.1016/j.elerap.2015.03.003>
- Mathayomchan, B., Taecharungroj, V., Wattanacharoensil, W., 2022. Evolution of COVID-19 tweets about Southeast Asian Countries: topic modelling and sentiment analyses. *Place Brand. Public Dipl.* <https://doi.org/10.1057/s41254-022-00271-5>
- McKinney, W., 2010. Data Structures for Statistical Computing in Python. *Proc. 9th Python Sci. Conf.* 56–61. <https://doi.org/10.25080/Majora-92bf1922-00a>
- Mediatama, G., 2022. Pengguna Aplikasi Dompot Digital Capai 87% [WWW Document]. [kontan.co.id](https://keuangan.kontan.co.id/news/pengguna-aplikasi-dompot-digital-capai-87). URL <https://keuangan.kontan.co.id/news/pengguna-aplikasi-dompot-digital-capai-87> (accessed 3.6.23).
- Mifrah, S., Benlahmar, E.H., 2020. Topic Modeling Coherence: A Comparative Study between LDA and NMF Models using COVID'19 Corpus. *Int. J. Adv. Trends Comput. Sci. Eng.* <https://doi.org/10.30534/ijatcse/2020/231942020>
- Mimno, D., Wallach, H., Talley, E., Leenders, M., McCallum, A., 2011. Optimizing Semantic Coherence in Topic Models, in: *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*. Presented at the EMNLP 2011, Association for Computational Linguistics, Edinburgh, Scotland, UK., pp. 262–272.
- Mohamed Ridhwan, K., Hargreaves, C.A., 2021. Leveraging Twitter data to understand public sentiment for the COVID-19 outbreak in Singapore. *Int. J. Inf. Manag. Data Insights* 1, 100021. <https://doi.org/10.1016/j.jjime.2021.100021>
- Monavia, R., 2022. Pengguna Twitter di Indonesia Capai 18,45 Juta pada 2022 [WWW Document]. [Dataindonesia.id](https://dataindonesia.id/digital/detail/pengguna-twitter-di-indonesia-capai-1845-juta-pada-2022). URL <https://dataindonesia.id/digital/detail/pengguna-twitter-di-indonesia-capai-1845-juta-pada-2022> (accessed 1.15.23).
- M.Sc, D.K., 2022. *Pengenalan Machine Learning dengan Python*. Elex Media Komputindo.
- Musfiroh, D., Khaira, U., Eko, P., Utomo, P., Suratno, T., 2021. Sentiment Analysis of Online Lectures in Indonesia from Twitter Dataset Using InSet Lexicon Analisis Sentimen terhadap Perkuliahan Daring di Indonesia dari Twitter Dataset Menggunakan InSet Lexicon. *MALCOM Indones. J. Mach. Learn. Comput. Sci.* 1, 24–33.

- Nasution, R.A., 2020. Analisis Persepsi Pedagang Pada Penggunaan QRIS Sebagai Alat Transaksi Umkm di Kota Medan (skripsi). Universitas Islam Negeri Sumatera Utara Medan.
- Natalina, S.A., Zunaidi, A., Rahmah, R., 2021. QUICK RESPONSE CODE INDONESIA STANDARD (QRIS) SEBAGAI STRATEGI SURVIVE USAHA MIKRO KECIL DAN MENENGAH (UMKM) DI MASA PANDEMI DI KOTA KEDIRI. *Istithmar J. Studi Ekon. Syariah* 5. <https://doi.org/10.30762/istithmar.v5i2.85>
- Nelli, F., 2018. *Python Data Analytics: With Pandas, NumPy, and Matplotlib*. Apress, Berkeley, CA. <https://doi.org/10.1007/978-1-4842-3913-1>
- Newman, D., Lau, J.H., Grieser, K., Baldwin, T., 2010a. Automatic evaluation of topic coherence, in: *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, HLT '10*. Association for Computational Linguistics, USA, pp. 100–108.
- Newman, D., Noh, Y., Talley, E., Karimi, S., Baldwin, T., 2010b. Evaluating topic models for digital libraries, in: *Proceedings of the 10th Annual Joint Conference on Digital Libraries, JCDL '10*. Association for Computing Machinery, New York, NY, USA, pp. 215–224. <https://doi.org/10.1145/1816123.1816156>
- Noor, S., Tajik, O., Golzar, J., 2022. Simple Random Sampling. *Int. J. Educ. Amp Lang. Stud.* 1, 78–82. <https://doi.org/10.22034/ijels.2022.162982>
- Noviansah, W., n.d. Duit dari QRIS Palsu Masuk Rekening Iman Mahlil, Bukan “Restorasi Masjid” [WWW Document]. *detiknews*. URL <https://news.detik.com/berita/d-6668548/duit-dari-qr-is-palsu-masuk-rekening-iman-mahlil-bukan-restorasi-masjid> (accessed 4.22.23).
- Nurkasanah, A., Hayaty, M., 2022. Feature Extraction using Lexicon on the Emotion Recognition Dataset of Indonesian Text. *Ultim. J. Tek. Inform.* 14, 20–27.
- Pal, A., Herath, T., De', R., Raghav Rao, H., 2021. Why do people use mobile payment technologies and why would they continue? An examination and implications from India. *Res. Policy* 50, 104228. <https://doi.org/10.1016/j.respol.2021.104228>
- Prayitno, M.A., Fadly, W., 2022. Pelatihan Pemanfaatan dan Pendampingan Pembuatan QRIS (QR Code Indonesian Standard) Sebagai Media Digitalisasi ZIS di Desa Glinggang Kabupaten Ponorogo.
- Priyanto, A., Ma'arif, M.R., 2018. Implementasi Web Scrapping dan Text Mining untuk Akuisisi dan Kategorisasi Informasi dari Internet (Studi Kasus: Tutorial Hidroponik). *Indones. J. Inf. Syst.* 1, 25–33. <https://doi.org/10.24002/ijis.v1i1.1664>
- Probyto Data Science and Consulting Pvt, L., 2020. *Data Science for Business Professionals: A Practical Guide for Beginners (English Edition)*. BPB Publications.
- Putra, I., 2017. Analisis topik informasi publik media sosial di surabaya menggunakan pemodelan latent dirichlet allocation (LDA).
- Putri, D.S., Ridwan, T., 2023. Analisis Sentimen Ulasan Aplikasi Pospay dengan Algoritma Support Vector Machine. *J. Ilm. Inform.* 11, 32–40.

- QRIS.id, 2023. QRIS (Quick Response Code Indonesian Standard) Satu QR Code untuk semua Payment [WWW Document]. URL <https://qris.id/homepage/> (accessed 2.2.23).
- Rachmat, A., Lukito, Y., 2016. SENTIPOL: DATASET SENTIMEN KOMENTAR PADA KAMPANYE PEMILU PRESIDEN INDONESIA 2014 DARI FACEBOOK PAGE.
- Řehůřek, R., Sojka, P., 2010. Software Framework for Topic Modelling with Large Corpora. Proceedings of LREC 2010 workshop New Challenges for NLP Frameworks.
- Rintyarna, B., 2021. Mapping acceptance of Indonesian organic food consumption under COVID-19 pandemic using sentiment analysis of Twitter dataset. *J. Theor. Appl. Inf. Technol.* 99, 1009–1019.
- Röder, M., Both, A., Hinneburg, A., 2015. Exploring the Space of Topic Coherence Measures, in: Proceedings of the Eighth ACM International Conference on Web Search and Data Mining, WSDM '15. Association for Computing Machinery, New York, NY, USA, pp. 399–408. <https://doi.org/10.1145/2684822.2685324>
- Rodrigues, A.P., Fernandes, R., Shetty, A., Lakshmana, K., Shafi, R.M., 2022. Real-time twitter spam detection and sentiment analysis using machine learning and deep learning techniques. *Comput. Intell. Neurosci.* 2022.
- Rusnawati, R., Farild, M., Ms, E.I., 2022. The Fintech E-Payment: The Impact to Financial Behavior. *LAA MAISYIR J. Ekon. Islam* 9, 20–32. <https://doi.org/10.24252/lamaisyir.v9i1.23793>
- Sandi, M.R., 2023. Permudah Transaksi, Seluruh Pasar Tradisional di Depok Sudah Terfasilitasi QRIS [WWW Document]. <https://www.idxchannel.com/>. URL <https://www.idxchannel.com/economics/permudah-transaksi-seluruh-pasar-tradisional-di-depok-sudah-terfasilitasi-qris> (accessed 3.31.23).
- Saputri, O.B., 2020. Preferensi konsumen dalam menggunakan quick response code Indonesia standard (qris) sebagai alat pembayaran digital. *KINERJA* 17, 237–247. <https://doi.org/10.30872/jkin.v17i2.7355>
- Saraswati, N.W.S., 2013. Naïve Bayes Classifier Dan Support Vector Machines Untuk Sentiment Analysis. *SESINDO 2013* 2013.
- Sari, F.V., Wibowo, A., 2019. ANALISIS SENTIMEN PELANGGAN TOKO ONLINE JD.ID MENGGUNAKAN METODE NAÏVE BAYES CLASSIFIER BERBASIS KONVERSI IKON EMOSI. *Simetris J. Tek. Mesin Elektro Dan Ilmu Komput.* 10, 681–686. <https://doi.org/10.24176/simet.v10i2.3487>
- Sari, W.A.N., Purnomo, H.D., 2022. Topic Modeling Using The Latent Dirichlet Allocation Method on Wikipedia Pandemic COVID-19 Data In Indonesia. *J. Tek. Inform. Jutif* 3, 1223–1230. <https://doi.org/10.20884/1.jutif.2022.3.5.321>
- Sarkar, D., 2016. Text Analytics with Python. <https://doi.org/10.1007/978-1-4842-2388-8>
- Sastrawi, 2023.
- Sbalchiero, S., Eder, M., 2020. Topic modeling, long texts and the best number of topics. Some Problems and solutions. *Qual. Quant.* 54, 1095–1108. <https://doi.org/10.1007/s11135-020-00976-w>

- Schmiedel, T., Müller, O., Brocke, J. vom, 2018. Topic Modeling as a Strategy of Inquiry in Organizational Research: A Tutorial With an Application Example on Organizational Culture. *Organ. Res. Methods* (in press). <https://doi.org/10.1177/1094428118773858>
- Septian, J.A., Fachrudin, T.M., Nugroho, A., 2019. Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF-IDF dan K-Nearest Neighbor. *INSYST J. Intell. Syst. Comput.* 1, 43–49. <https://doi.org/10.52985/insyst.v1i1.36>
- Shihab, M.Q., 2000. Tafsir Al-Mishbah: pesan, kesan dan keserasian Al-Qur'an. Lentera Hati.
- Shurrab, S., Shannak, Y., Almshnanah, A., Khazaleh, H., Najadat, H., 2021. Attitudes Evaluation Toward COVID-19 Pandemic: An Application of Twitter Sentiment Analysis and Latent Dirichlet Allocation, in: 2021 12th International Conference on Information and Communication Systems (ICICS). Presented at the 2021 12th International Conference on Information and Communication Systems (ICICS), pp. 265–272. <https://doi.org/10.1109/ICICS52457.2021.9464558>
- Singh, A.S., Masuku, M.B., 2014. Sampling techniques & determination of sample size in applied statistics research: An overview. *Int. J. Econ. Commer. Manag.* 2, 1–22.
- Statista, 2022. Countries with most Twitter users 2022 [WWW Document]. Statista. URL <https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/> (accessed 1.15.23).
- TimTVOne, 2022. QRIS Bisa Digunakan di Thailand, Transaksi jadi Lebih Mudah [WWW Document]. URL <https://www.tvonenews.com/ekonomi/63678-qr-is-bisa-digunakan-di-thailand-transaksi-jadi-lebih-mudah> (accessed 1.14.23).
- Tobing, G.J., Abubakar, L., Handayani, T., 2021. Analisis Peraturan Penggunaan QRIS Sebagai Kanal Pembayaran Pada Praktik UMKM Dalam Rangka Mendorong Perkembangan Ekonomi Digital. *Acta Com. J. Huk. Kenotariatan* 6, 491–509.
- Vasant, P., Weber, G.-W., Marmolejo-Saucedo, J.A., Munapo, E., Thomas, J.J., 2022. Intelligent Computing & Optimization: Proceedings of the 5th International Conference on Intelligent Computing and Optimization 2022 (ICO2022). Springer Nature.
- Vayansky, I., Kumar, S.A.P., 2020. A review of topic modeling methods. *Inf. Syst.* 94, 101582. <https://doi.org/10.1016/j.is.2020.101582>
- Wedel, I., Palk, M., Voß, S., 2022. A Bilingual Comparison of Sentiment and Topics for a Product Event on Twitter. *Inf. Syst. Front.* 24, 1635–1646. <https://doi.org/10.1007/s10796-021-10169-x>
- Wettayaprasit, W., Laosen, N., Chevakidagarn, S., 2007. Data filtering technique for neural networks forecasting. Presented at the Proceedings of the 7th WSEAS International Conference on Simulation, Modelling and Optimization, World Scientific and Engineering Academy and Society (WSEAS), pp. 225–230.
- Xie, C., Gao, J., Tao, C., 2017. Big Data Validation Case Study, in: 2017 IEEE Third International Conference on Big Data Computing Service and Applications (BigDataService). Presented at the 2017 IEEE Third

- International Conference on Big Data Computing Service and Applications (BigDataService), pp. 281–286.
<https://doi.org/10.1109/BigDataService.2017.44>
- Yin, H., Song, X., Yang, S., Li, J., 2022. Sentiment analysis and topic modeling for COVID-19 vaccine discussions. *World Wide Web* 25, 1067–1083.
<https://doi.org/10.1007/s11280-022-01029-y>
- Yiran, Y., Srivastava, S., 2019. Aspect-based Sentiment Analysis on mobile phone reviews with LDA, in: Proceedings of the 2019 4th International Conference on Machine Learning Technologies, ICMLT 2019. Association for Computing Machinery, New York, NY, USA, pp. 101–105.
<https://doi.org/10.1145/3340997.3341012>
- Yovita, 2021. Perbedaan QRIS Statis dan Dinamis yang Kelihatan Sama - GoBiz. GoBiz - Pus. Pengetah. URL <https://gobiz.co.id/pusat-pengetahuan/perbedaan-qriskstatis-dan-dinamis/> (accessed 3.30.23).
- Yulivia, A.P., 2023. SENTIMENT ANALYSIS LAYANAN PESAN ANTAR MAKANAN PADA TWITTER MENGGUNAKAN INSET LEXICON (STUDI KASUS: GOFOOD DAN GRABFOOD) [WWW Document]. URL <http://digilib.unila.ac.id/69658/> (accessed 5.5.23).
- Zhai, C., Massung, S., 2016. Text Data Management and Analysis: A Practical Introduction to Information Retrieval and Text Mining. Association for Computing Machinery and Morgan & Claypool.
- Žižka, J., Dařena, F., Svoboda, A., 2019. Text Mining with Machine Learning: Principles and Techniques. CRC Press, Boca Raton.
<https://doi.org/10.1201/9780429469275>
- Zuhri, F., Alamsyah, A., 2017. Analisis Sentimen Masyarakat Terhadap Brand Smartfren Menggunakan Naive Bayes Classifier Di Forum Kaskus. *EProceedings Manag.* 4.

UIN SUNAN AMPEL
 S U R A B A Y A