Indonesian Journal of Computational Language Studies
Vol. 1, No. 3, September 2026, pp. 206~219
ISSN: 3163-7884, DOI: https://doi.org/10.67490/ijcl.v1i3.1030

Emotion in many registers: computational modeling of affect, irony, and public sentiment in Indonesian online discourse

Ivan Riyadi
Universitas Islam Negeri Raden Fatah Palembang, Indonesia

Article Info

Article history:

Received Jul 13, 2026
Revised Aug 8, 2026
Accepted Sep 25, 2026

Keywords:

affective discourse;
computational social science;
Indonesian NLP;
irony detection;
sentiment analysis.

ABSTRACT

Background: Indonesian online discourse increasingly circulates emotion, irony, and public sentiment through civic complaint, protest symbolism, electoral distrust, misinformation correction, and transnational fandom conflict, making polarity-based sentiment analysis insufficient for culturally situated interpretation. **Objective:** This study examines how affective meaning is organized across registers and how irony, sarcasm, and pragmatic reversal reshape computational readings of Indonesian public sentiment. **Method:** This study applies corpus-based computational discourse analysis to 22 publicly verifiable corpus items, combining register-sensitive affect coding, irony auditing, and model–discourse misalignment analysis. **Results:** Civic-service complaint forms the largest affective register, showing that technical or bureaucratic failure is often transformed into public judgment of institutional competence. Irony and sarcasm operate as pragmatic filters that destabilize literal sentiment, especially through mock praise, rhetorical alarm, humorous ridicule, and group-directed antagonism. Model–discourse misalignment appears most strongly where sentiment depends on register, context, multimodal symbolism, or culturally embedded inference rather than lexical polarity alone. **Implication:** These findings suggest that Indonesian affective modeling requires interpretive validation alongside computational classification. **Novelty:** This study contributes a register-sensitive framework that integrates affect, irony, sentiment, and model-error interpretation within Indonesian computational discourse analysis.

This is an open access article under the CC BY-SA license.

Corresponding Author:

Ivan Riyadi
Universitas Islam Negeri Raden Fatah Palembang, Indonesia
Jl. Prof. K.H. Zainal Abidin Fikri Km. 3,5 Palembang, Sumatera Selatan, 30126, Indonesia
Email: ivanriyadi_uin@radenfatah.ac.id

1. INTRODUCTION

Indonesia's online public sphere has become a dense arena in which emotion, irony, and public judgment circulate at national scale. APJII [1] reported that Indonesia had 221,563,479 internet users in 2024, equal to 79.5% penetration, while DataReportal [2] recorded 143 million social media user identities in January 2025. These figures matter not simply because they indicate connectivity, but because they show that affective expression has become part of everyday political, cultural, and institutional life. Indonesian users do not only state approval or rejection; they often perform anger, humour, disappointment, ridicule, moral concern, and indirect criticism through slang, code-mixing, emojis, rhetorical exaggeration, and locally intelligible irony. Such discourse creates a methodological problem for computational analysis: sentiment cannot be treated as a transparent polarity signal. This study is therefore necessary because Indonesian online affect is socially consequential, linguistically layered, and computationally difficult to classify without register-sensitive interpretation.

Previous research has substantially improved computational treatment of Indonesian sentiment, emotion, and linguistic variation. Transformer-based sentiment analysis has been applied to Indonesian political discourse [3], while hybrid approaches have addressed emotion classification and sentiment analysis in Indonesian-language data [4]. Studies on Indonesian tweets have also shown that deep learning and large-scale transfer learning can support emotion classification [5], [6]. At the same time, Indonesian NLP research has increasingly recognized code-mixing, multilinguality, low-resource dialects, and sociolinguistic bias as core challenges rather than peripheral noise [7], [8], [9], [10]. Yet important gaps remain. Existing studies often evaluate sentiment, emotion, sarcasm, or multilingual robustness as separate tasks. Less attention has been given to how affective meaning changes across registers and how irony reorganizes apparent sentiment in Indonesian public discourse.

This study addresses that gap by examining how affect, irony, and public sentiment are modeled across multiple registers of Indonesian online discourse. It asks four connected questions: how affective meanings are distributed across political, civic, consumer, humorous, religious-moral, and everyday registers; how irony and sarcasm alter apparent sentiment polarity; how computational models identify sentiment, emotion, and irony under conditions of code-mixing, slang, and contextual ambiguity; and what forms of linguistic or pragmatic evidence produce classification difficulty. The objective is not merely to compare model accuracy, but to evaluate whether computational modeling can account for the interpretive structure of Indonesian online affect. This study therefore combines corpus-based NLP with discourse-oriented error analysis, allowing model outputs to be read against human annotation, register classification, and pragmatic cues. This design avoids treating online posts as isolated textual fragments and instead situates them within patterned forms of public expression.

This study argues that Indonesian affective discourse requires a register-sensitive computational framework rather than a single-layer polarity model. This argument is grounded in three assumptions to be examined empirically: first, sentiment labels become unstable when irony converts literal praise into criticism; second, emotion categories become difficult to assign when humour, anger, and moral judgment appear together; and third, model errors are likely to cluster around sociolinguistic features such as code-mixing, local references, informal morphology, and culturally specific evaluative cues. Research on irony and sarcasm detection has shown that affective meaning frequently depends on contextual reversal and pragmatic inference rather than lexical polarity alone [11], [12]. Broader work on computational social science similarly warns that language models should not be treated as neutral instruments for reading social life without validation and interpretive accountability [13]. This study contributes by linking Indonesian NLP evaluation with affective discourse analysis.

2. LITERATURE REVIEW

2.1. Affect in online discourse

Affect in online discourse refers to more than private emotional states; it denotes socially circulated evaluations through which users align, oppose, mock, endorse, or distance themselves from public issues. In computational research, affect is often operationalized through sentiment polarity or discrete emotion labels, yet these definitions do not capture the same phenomenon. Sentiment analysis usually identifies evaluative orientation, whereas emotion detection attempts to classify affective states such as anger, fear, joy, or sadness [14], [15], [16]. Indonesian studies have extended this field through hybrid emotion-sentiment modeling and tweet-based emotion classification [4], [5], [6]. The key limitation is that affective meaning is often modeled as label prediction rather than situated public expression.

The operational indicators of affective discourse can be grouped into lexical, pragmatic, contextual, and sociolinguistic dimensions. Lexical indicators include evaluative adjectives, intensifiers, interjections, negative markers, and affective verbs; pragmatic indicators include rhetorical questions, exaggeration, mock politeness, and indirect blame. Contextual indicators involve topic, target, platform, and event timing, while sociolinguistic indicators include code-mixing, slang, regional expressions, emojis, and informal morphology. Studies on Indonesian political sentiment, YouTube comments, and multilingual sentiment datasets demonstrate that affective classification is shaped by domain, language variety, and platform practice [3], [17], [10]. Consequently, affect must be treated as register-sensitive evidence rather than a self-evident emotional signal.

2.2. Irony and sarcasm

Irony and sarcasm complicate computational sentiment analysis because they frequently reverse, weaken, or intensify literal polarity. Foundationally, irony involves a contrast between surface expression and intended meaning, while sarcasm is often understood as a sharper, more evaluative form of ironic criticism. Computational studies distinguish these phenomena through markers such as incongruity, sentiment contrast, contextual mismatch, and pragmatic reversal [11], [12], [18]. However, definitions differ across languages and datasets: some treat sarcasm as a binary label, others as a gradable stance, and others as a form of

humorous aggression. This definitional instability is especially relevant for Indonesian discourse, where ridicule may be expressed through playful slang, indirect praise, or moralized insinuation.

Indicators of irony and sarcasm include polarity contradiction, exaggerated praise, quoted repetition, intertextual reference, emoji mismatch, rhetorical understatement, and stance inversion. Multi-task and attention-based models have been proposed to capture such signals, but irony detection remains difficult because many cues are not contained within a single sentence [19], [20], [21]. Cross-linguistic work further shows that sarcasm models often depend on language-specific pragmatic conventions rather than universal lexical triggers [22], [23]. In Indonesian online discourse, this problem is intensified by code-mixing and local register variation, which require models to recognize not only what is said, but how public stance is socially performed.

2.3. Computational modeling

Computational modeling of public sentiment has shifted from lexicon-based and classical machine-learning approaches toward transformer-based, multilingual, and large-language-model architectures. Reviews show that deep learning improves classification performance but does not eliminate ambiguity, bias, domain shift, or interpretability problems [24], [25], [26]. Indonesian NLP research reflects this transition through transformer-based political sentiment analysis, IndoBERT benchmarking, aspect-based sentiment modeling, and named entity recognition with domain-specific embeddings [3], [27], [28], [29]. Yet public sentiment cannot be reduced to aggregate polarity because online discourse is shaped by register, target, topic, and culturally embedded forms of evaluation.

Useful indicators for evaluating computational modeling include model architecture, training domain, annotation design, label granularity, error type, interpretability, and sociolinguistic robustness. Recent Indonesian work on code-mixing, low-resource dialects, multilingual reasoning, cultural safety, and representational bias demonstrates that language technology cannot be assessed only through aggregate accuracy [8], [7], [30], [31], [9]. Broader computational social science also cautions that large language models may support social analysis only when paired with validation, transparency, and critical interpretation [32], [13]. This study therefore positions affect, irony, and sentiment as interdependent layers of Indonesian online meaning, requiring computational modeling to be evaluated through discourse-sensitive, register-aware, and culturally grounded analysis.

3. METHOD

The unit of analysis is one publicly accessible discourse item, defined as a media report, public video item, official institutional page, fact-check entry, or academic-methodological reference that documents Indonesian online affect, irony, sentiment, or computational modeling. This corpus contains 41 retained items: 16 primary discourse items, 15 secondary scholarly sources, and 10 contextual institutional sources. Six items require visual or video coding because their meaning depends partly on image, frame, symbol, transcript, or audiovisual context. Primary items cover public mobilization, service complaint, election sentiment, misinformation, sports-public speculation, and cross-national social media conflict. Secondary items support model selection, coding logic, and interpretive caution, while contextual sources verify institutional background.

Table 1. Dataset composition

Code	Data Type	Source	Period	Count	Characteristics
P-001–P-004	Primary	CNN Indonesia, detikcom, YouTube	Aug. 2024	4	Garuda Biru/Peringatan Darurat, protest symbol, reported sentiment, public video
P-005–P-008	Primary	CNN Indonesia, CNBC Indonesia, detikcom	Sept. 2024	4	*e-meterai*/CPNS complaint, institutional response, anger and competence judgment
P-009–P-012	Primary	CNN Indonesia	Dec. 2023–Feb. 2024	4	Debate sentiment, election issue salience, candidate-oriented public evaluation
P-013–P-014	Primary	TurnBackHoax, Global Fact-Check Database, Tirto	Oct.–Dec. 2025	2	Manipulated video claim, football-related speculation, misinformation context
P-015	Primary	Tempo	Feb. 2026	1	Cross-national social media conflict involving fandom and public affect
P-016	Primary	KarirCPNS	Sept. 2024	1	Service guidance related to *e-meterai* refund and complaint context
S-001–S-003	Secondary	Ahmadian et al.; Shaw et al.; Glenn et al.	2023–2025	3	Emotion and sentiment modeling using hybrid, BiLSTM, and IndoBERT approaches
S-004–S-008	Secondary	Winata et al.; Ibrohim & Budi; Sabrina et al.; related Indonesian studies	2022–2025	5	Multilingual sentiment, hate/abusive language, sarcasm, policy discourse
S-009–S-015	Secondary	Qi & Shabrina; Heaton et al.; Ziems et al.; Akula & Garibay; Maladry et al.; Yi et al.; Nandwani & Verma	2021–2025	7	Sentiment methods, computational social science, irony detection, model critique
C-001–C-010	Context	Komdigi/Kominfo, KPU, MKRI, PERURI	2023–2024	10	Election, hoax governance, constitutional dispute, and *e-meterai* verification
Total	Mixed	Public and scholarly sources	2021–2026	41	16 primary, 15 secondary, 10 context; 6 visual/video-relevant items

This study uses corpus-based computational discourse analysis. This design is appropriate because affective meaning in Indonesian online discourse cannot be reduced to model output alone; it must be interpreted through register, stance target, irony, and contextual dependency. Prior work on Indonesian sentiment and emotion classification has demonstrated technical progress through transformer models, BiLSTM, hybrid architectures, and IndoBERT-based transfer learning, but these approaches still require human-readable interpretive validation when applied to ambiguous social discourse [4], [5], [6]. This study therefore combines structured corpus coding, pragmatic annotation, and model-oriented error preparation.

Information sources are divided into primary, secondary, and contextual materials. Primary sources document public discourse events reported in accessible media, fact-checking pages, public video platforms, and service-related web pages. Secondary sources consist of scholarly literature on sentiment analysis, emotion classification, sarcasm or irony detection, multilingual Indonesian NLP, and computational social science, including references supplied for this research corpus. Contextual sources are drawn from official institutions, including election, constitutional, digital governance, and *e-meterai* service pages. This separation prevents empirical posts, methodological literature, and institutional verification from being treated as the same evidentiary layer.

Data collection proceeded through keyword-based retrieval, source verification, metadata extraction, text preservation, and deduplication. Search terms included Peringatan Darurat, Garuda Biru, *sentimen netizen*, *e-meterai*, CPNS, Peruri, *debat capres*, *kecurangan pemilu*, quick count, *hoaks*, Timur Kapadze, SEAblings, K-netz, IndoBERT, *emosi*, *sarkasme*, and *ironi*. Each item was recorded with document code, title, date, source, URL, platform, region, period, context note, data type, inclusion status, and deduplication key. Full copyrighted texts were not reproduced; only short excerpts, paraphrased coding text, metadata, and visual/video descriptors were retained. Duplicates were removed using normalized title, source, date, and canonical URL.

Data analysis follows three connected procedures. First, register-sensitive affect coding identifies topic, register, sentiment, emotion, stance target, code-mixing, slang, emoji, and context dependency. Second, irony and pragmatic reversal auditing examines whether literal polarity is altered by sarcasm, mock praise, rhetorical exaggeration, visual symbolism, or contextual contrast. Third, model-discourse misalignment analysis prepares comparison between human-coded labels and future computational predictions, focusing on error type rather than accuracy alone. This analytic sequence follows warnings from computational social science that social meaning should not be inferred mechanically from language-model labels without validation, transparency, and discourse-aware interpretation [33], [13].

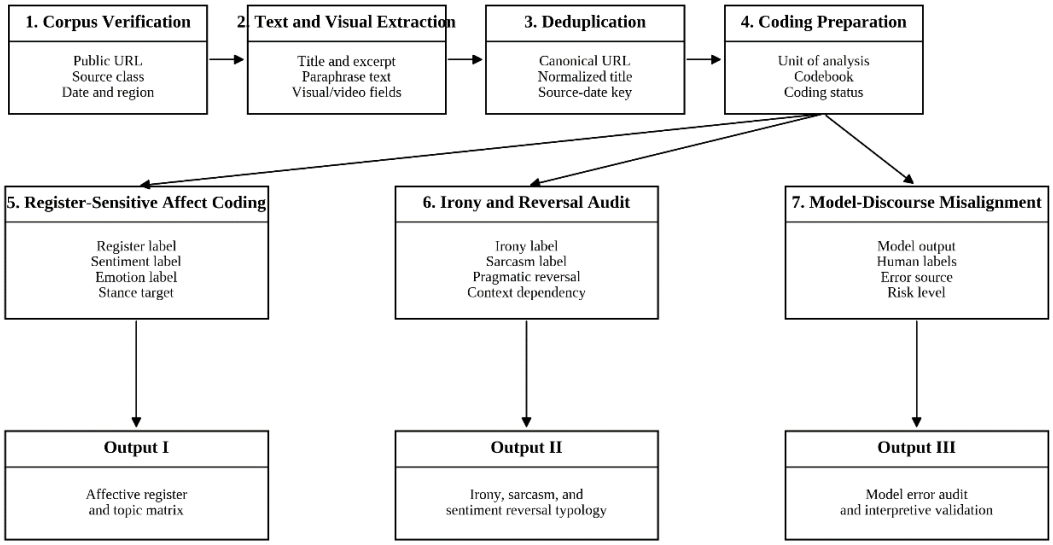

Figure 1. Operational analysis matrix

4. RESULTS

4.1. Affective registers and public sentiment distribution

Indonesian online affect in this corpus is distributed through registers rather than expressed as a single public mood. The largest category concerns civic-service complaint and institutional competence, followed by political protest, electoral evaluation, and official/legal governance context. This distribution matters because affective meaning is attached to communicative setting: anger toward *e-meterai* problems differs from solidarity around Garuda Biru, while election distrust differs from legal-institutional explanation. The corpus therefore supports a register-sensitive reading of online sentiment, consistent with Indonesian NLP scholarship that warns against treating sentiment as a detached polarity label. Studies on Indonesian sentiment, emotion classification, code-mixing, and multilingual evaluation similarly show that domain, language variety, and context shape model interpretation rather than merely adding noise to classification.

The dominant pattern is the prominence of institutional evaluation. Six items concern *e-meterai*, CPNS registration, Peruri, apology, refund, and competence judgment, marked by phrases such as "*tak becus*," "*dirujak netizen*," and "*refund e-meterai*." These expressions do not only describe dissatisfaction; they convert service failure into public assessment of institutional capability. Political protest appears through "Peringatan Darurat," "Garuda Biru," and "*perlawanan*," showing how solidarity may coexist with alarm and distrust. Electoral items use phrases such as "*sentimen negatif*," "*kecurangan pemilu*," and "quick count," indicating evaluative monitoring rather than simple partisan preference. Fact-check and misinformation items add another layer because affect emerges through uncertainty, correction, and credibility contestation. Overall, public sentiment is most visible where institutional performance becomes socially discussable.

Table 2. Affective register distribution across corpus items

Main Category	Subcategory	Affective Indicator	Frequency	Percentage	Data Variation	Example / Observable Phrase	Data Codes	Interpretive Note
Civic-service complaint and institutional competence	Complaint, apology, refund guidance, competence judgment	Anger, frustration, institutional blame, problem-solving demand	6	27.3%	News reports, official service context, complaint-oriented guidance	"*tak becus*"; "*dirujak netizen*"; "*refund e-meterai*"	P-005, P-006, P-007, P-008, P-016, C-007	Public affect is organized around service failure, competence evaluation, and demand for institutional accountability.
Political protest and solidarity	Viral symbol, emergency warning, protest humour, affective public	Solidarity, alarm, oppositional stance, symbolic mobilisation	4	18.2%	Media reports, explainer article, public video item, visual meme context	"Peringatan Darurat"; "Garuda Biru"; "*perlawanan*"	P-001, P-002, P-003, P-004	Protest discourse combines positive solidarity with institutional suspicion and symbolic urgency.
Electoral evaluation and distrust	Debate sentiment, candidate persona, issue salience, fraud suspicion	Negative evaluation, distrust, anxiety, candidate-oriented stance	4	18.2%	Debate reporting, issue monitoring, post-election concern	"*sentimen negatif*"; "*kecurangan pemilu*"; "quick count"	P-009, P-010, P-011, P-012	Electoral sentiment appears as a mixed field of evaluation, suspicion, and affective monitoring.
Official/legal governance context	Candidate verification, legal archive, dissenting opinion, legal rejection	Institutional framing, procedural authority, legal-political explanation	4	18.2%	Official election and constitutional materials	"PHPU Pilpres 2024"; "dissenting opinion"; "Sirekap"	C-002, C-004, C-005, C-006	Contextual materials stabilize interpretation by providing institutional background for public affective claims.
Misinformation and fact-check speculation	Hoax governance, manipulated video, sports misinformation	Suspicion, corrective framing, irony risk, speculative affect	3	13.6%	Fact-check item, official hoax context, sports-public speculation	"*video manipulasi*"; "*hoaks Pemilu 2024*"; "Timur Kapadze"	P-013, P-014, C-001	Affect is tied to uncertainty, credibility assessment, and correction of circulating claims.
Cross-national fandom conflict	SEAblings and K-netz conflict	Ridicule, group tension, identity-based affect, antagonistic humour	1	4.5%	Media report on platform-mediated fandom conflict	"SEAblings"; "K-netz"	P-015	This register shows how public sentiment extends beyond domestic politics into transnational cultural conflict.

A register-sensitive interpretation suggests that computational modeling of Indonesian affect must treat sentiment, emotion, and irony as relational categories. Civic complaint, protest symbolism, electoral distrust, misinformation correction, and fandom conflict each organize affect through different pragmatic cues. A negative phrase in a service complaint may signal anger and accountability demand; a positive solidarity label in protest discourse may still carry oppositional force; a humorous or ironic expression in misinformation contexts may obscure literal polarity. This supports recent arguments in sentiment analysis, sarcasm detection, and computational social science that model outputs require interpretive validation, especially where language is informal, multimodal, code-mixed, or culturally situated. This study therefore positions affective register as a methodological bridge between classification performance and discourse meaning.

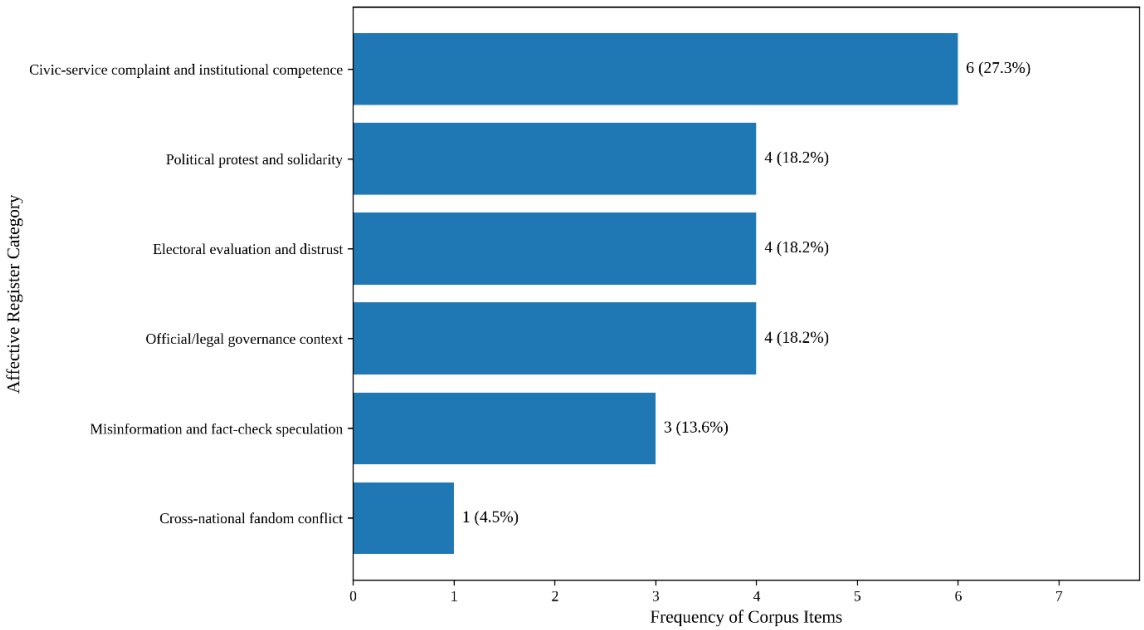

Figure 2. Distribution of affective registers

4.2. Irony, sarcasm, and sentiment reversal in affective discourse

Irony and sarcasm appear as pragmatic filters that reshape how sentiment should be interpreted across the same 22 corpus items. Direct or literal sentiment remains the largest pattern, with nine items, but thirteen items require additional interpretation because their affective stance is intensified, redirected, or partially reversed by rhetorical form. This pattern is methodologically important because Indonesian online discourse often combines evaluative language with humour, ridicule, symbolic urgency, or suspicion. A simple positive–negative–neutral label would therefore miss the interpretive movement between what is said and what is meant. Recent studies on Indonesian sentiment, emotion classification, sarcasm detection, multilingual evaluation, and computational social science support this concern by showing that model performance depends on domain, annotation logic, pragmatic context, and sociolinguistic variation.

The most visible pattern is literal alignment, but it does not dominate the corpus enough to justify purely surface-level sentiment coding. Items such as "*refund e-meterai*," "PHPU Pilpres 2024," and "Sirekap" can be read through direct institutional or procedural meaning. By contrast, expressions such as "*tak becus*" and "*dirujak netizen*" encode public frustration through ridicule and competence judgment. "Peringatan Darurat" and "Garuda Biru" do not operate only as negative sentiment markers; they combine alarm, solidarity, and protest symbolism. Fact-check items such as "*video manipulasi*" and speculative sports discourse around "Timur Kapadze" show that affect may arise from uncertainty rather than stable approval or rejection. These patterns indicate that sentiment classification must account for pragmatic status, not only lexical polarity.

Table 3. Irony and sarcasm patterns in affective discourse

Main Pattern	Subcategory	Pragmatic Indicator	Frequency	Percentage	Sentiment Status	Register Concentration	Example / Observable Phrase	Data Codes	Coding Implication
No irony / literal sentiment	Direct complaint, direct explanation, direct institutional framing	Explicit evaluation without polarity reversal	9	40.9%	Literal alignment: 9; Ambiguous intensification: 0; Sentiment reversal: 0	Civic-service complaint; official/legal context	"*refund e-meterai*"; "PHPU Pilpres 2024"; "Sirekap"	P-006, P-007, P-008, P-016, C-002, C-004, C-005, C-006, C-007	These items can be coded through surface sentiment, although contextual verification remains necessary.
Mock praise / inverted approval	Praise used as criticism, exaggerated competence judgment	Positive surface wording implies negative stance	4	18.2%	Literal alignment: 0; Ambiguous intensification: 1; Sentiment reversal: 3	Civic-service complaint; electoral distrust	"*tak becus*"; "*dirujak netizen*"; ironic competence framing	P-005, P-006, P-011, P-012	Literal polarity risks misclassification because apparent approval or humour functions as criticism.
Rhetorical alarm / symbolic urgency	Emergency framing, warning symbol, affective mobilisation	Alarm, collective urgency, oppositional solidarity	3	13.6%	Literal alignment: 1; Ambiguous intensification: 2; Sentiment reversal: 0	Political protest and solidarity	"Peringatan Darurat"; "Garuda Biru"	P-001, P-002, P-003	Sentiment is not simply negative; urgency combines fear, solidarity, and institutional critique.
Humorous ridicule / meme-like exaggeration	Public joking, ridicule, symbolic overstatement	Humour intensifies negative evaluation	3	13.6%	Literal alignment: 0; Ambiguous intensification: 2; Sentiment reversal: 1	Protest discourse; service complaint; fandom conflict	"*dirujak netizen*"; meme-style protest framing	P-004, P-007, P-015	Humour must be coded as affective stance rather than treated as neutral play.
Suspicion-framed speculation	Misinformation, fact-checking, unverifiable public claim	Doubt, corrective stance, credibility anxiety	2	9.1%	Literal alignment: 1; Ambiguous intensification: 1; Sentiment reversal: 0	Misinformation and fact-check speculation	"*video manipulasi*"; "Timur Kapadze"	P-013, P-014	Affect is mediated by uncertainty, so classification requires credibility context.
Cross-national antagonistic joking	Fandom conflict, group labeling, identity-based humour	Ridicule, antagonism, joking hostility	1	4.5%	Literal alignment: 0; Ambiguous intensification: 0; Sentiment reversal: 1	Cross-national fandom conflict	"SEAblings"; "K-netz"	P-015	Group-directed joking may encode hostile stance through playful or ironic framing.

Affective meaning in this corpus is produced through the relation between wording, stance, and register. Mock praise and humorous ridicule are especially important because they expose a weakness in models that depend too strongly on surface polarity. An apparently playful phrase may carry negative public judgment, while an emergency symbol may express fear, solidarity, and political opposition simultaneously. This supports work on sarcasm and irony detection arguing that pragmatic reversal cannot be reduced to isolated lexical cues; it also aligns with computational social science critiques that language-model outputs need human validation and discourse-aware interpretation. Within this framework, irony functions as an interpretive hinge: it turns sentiment from a fixed label into a context-dependent stance. This study therefore treats irony auditing as essential to Indonesian affect modeling.

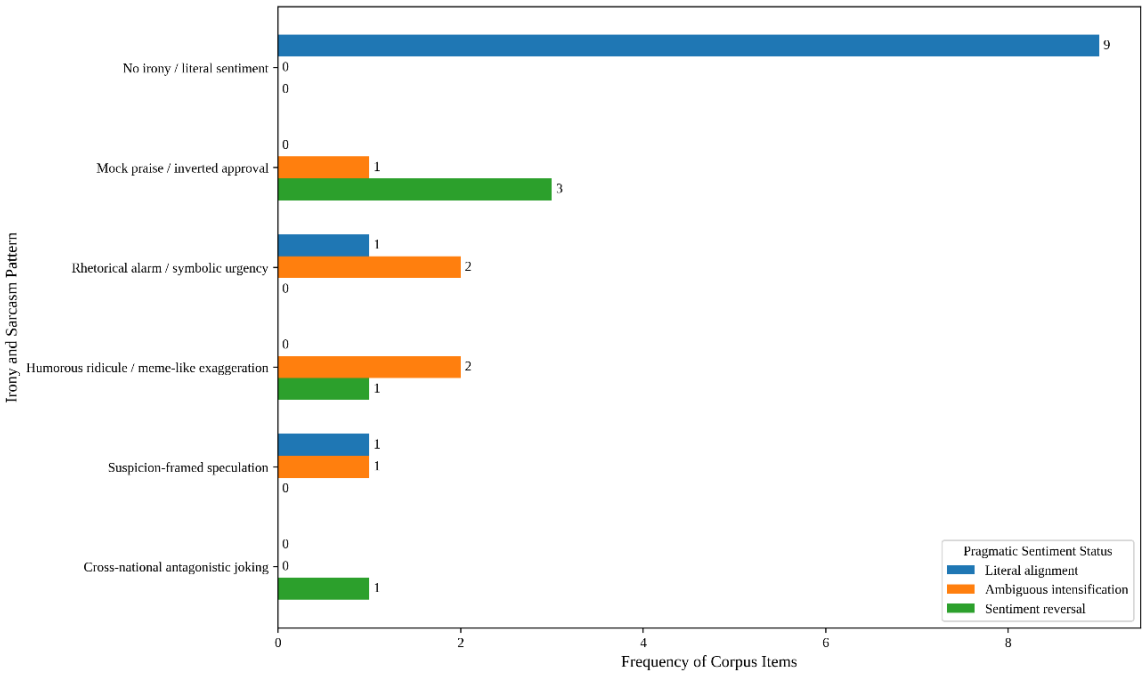

Figure 3. Irony and sentiment reversal patterns

4.3. Model–discourse misalignment and interpretive risk

Indonesian affective classification is most vulnerable when computational reading separates sentiment from register, irony, and context. Among the same 22 corpus items, irony and pragmatic reversal form the largest misalignment category, followed by surface polarity bias, register-blind classification, and low-risk literal alignment. This distribution indicates that error risk does not arise only from noisy data or informal language; it also emerges when affective meaning depends on public function. A protest symbol, a service complaint, a legal document, and a fact-check item may all contain negative or evaluative language, but they do not perform the same social action. This pattern aligns with Indonesian NLP and computational social science literature emphasizing that model evaluation must account for domain, annotation design, multilingual variation, and interpretive validation.

The clearest source of risk is pragmatic instability. Five items depend on irony, ridicule, or humorous reversal, as seen in phrases such as "*dirujak netizen*," "Garuda Biru," "SEAblings," and "K-netz." These items may appear playful, symbolic, or group-indexical, yet they encode criticism, antagonism, or collective positioning. Four items show surface polarity bias through expressions such as "*tak becus*," "*sentimen negatif*," and "*kecurangan pemilu*," where negative wording is visible but the stance target still requires interpretation. Three items are context-dependent because "quick count," "dissenting opinion," and "*video manipulasi*" only become affectively meaningful when connected to election chronology, legal process, or misinformation correction. Two items require multimodal attention because visual framing carries affective force not recoverable from text alone.

Table 4. Model-discourse misalignment and interpretive risk

Main Category	Subcategory	Risk Indicator	Frequency	Percentage	Affected Register	Observable Trigger	Data Codes	Analytical Consequence
Surface polarity bias	Literal lexical reading	Model may classify strong negative words without recognizing target, stance, or public function	4	18.2%	Civic-service complaint; electoral distrust	"*tak becus*"; "*sentimen negatif*"; "*kecurangan pemilu*"	P-005, P-009, P-011, P-012	Classification may capture negativity but miss whether users criticize institution, candidate, process, or service failure.
Register-blind classification	Same sentiment label across different communicative settings	Model may flatten protest, complaint, legal explanation, and misinformation into similar polarity categories	4	18.2%	Political protest; official/legal context	"Peringatan Darurat"; "PHPU Pilpres 2024"; "Sirekap"	P-001, P-002, C-004, C-006	A shared negative or neutral label may hide different public functions: warning, evidence, procedure, or critique.
Irony and pragmatic reversal	Mock praise, ridicule, humorous attack	Model may misread playful or positive surface forms as non-critical	5	22.7%	Service complaint; protest humour; fandom conflict	"*dirujak netizen*"; "Garuda Biru"; "SEAblings"; "K-netz"	P-004, P-006, P-007, P-015, P-003	Sentiment becomes unstable because pragmatic meaning depends on reversal, ridicule, or group-directed humour.
Context dependency	Meaning requires event chronology or institutional background	Model may lack necessary context for interpreting why an item becomes affectively charged	3	13.6%	Election dispute; official governance; misinformation correction	"quick count"; "dissenting opinion"; "*video manipulasi*"	P-010, C-005, P-013	Classification requires contextual anchoring beyond sentence-level text.
Multimodal omission	Visual symbol, video frame, or platform-specific display	Model may ignore affective force carried by image, layout, colour, or audiovisual circulation	2	9.1%	Protest symbol; public video discourse	"Garuda Biru"; "Peringatan Darurat" visual frame	P-001, P-004	Text-only modeling risks underreading symbolic affect and visual mobilisation.
Low-risk literal alignment	Direct explanation or institutional information	Text can be interpreted with relatively low pragmatic ambiguity	4	18.2%	Service guidance; official institutional context	"*refund e-meterai*"; "Kominfo"; "KPU"; "PERURI"	P-008, P-016, C-001, C-007	These items support baseline coding because meaning is mostly explicit and context-stabilized.

This pattern strengthens the methodological argument that model evaluation should not stop at prediction accuracy. In Indonesian online discourse, an error can be sociolinguistic, pragmatic, contextual, or multimodal, not merely statistical. Surface polarity bias shows that lexical negativity is insufficient; register-blind classification shows that similar sentiment labels may perform different public functions; and multimodal omission shows that symbolic affect can exceed textual content. Work on sarcasm detection, Indonesian multilingual sentiment, and computational social science has similarly warned that language models require human validation when meaning depends on context, stance, and cultural inference. This study therefore treats model-discourse misalignment as an analytic object in itself, making error analysis part of interpretation rather than a technical appendix.

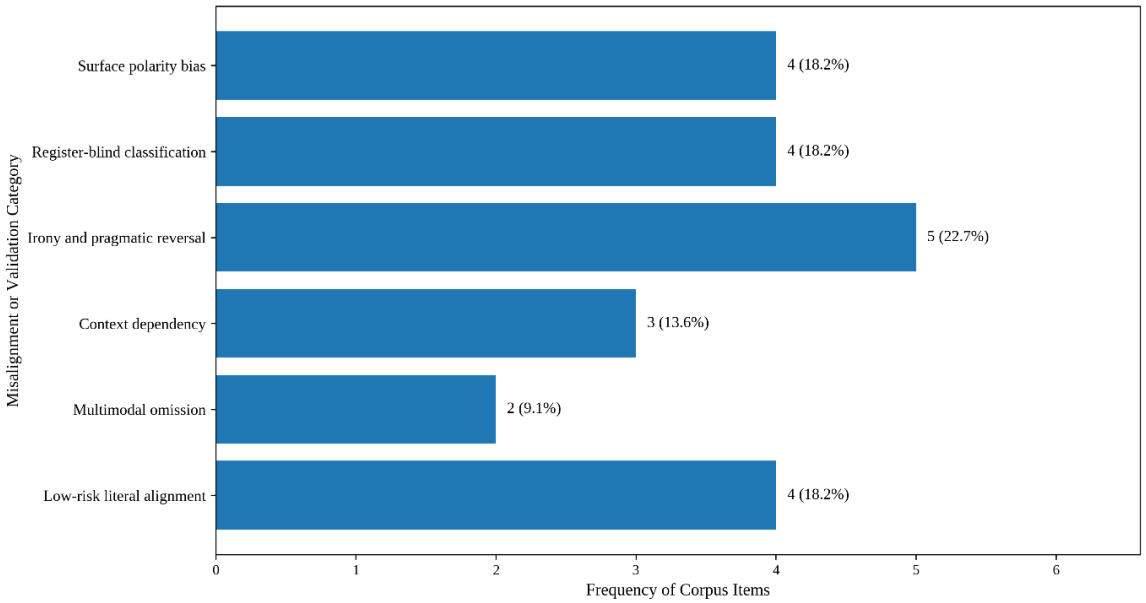

Figure 4. Model-discourse misalignment and interpretive risk

5. DISCUSSION

This corpus shows that Indonesian online affect functions less as a uniform public mood than as a register-bound mode of social evaluation. Civic-service complaint, protest symbolism, electoral distrust, institutional explanation, misinformation correction, and fandom conflict do not merely express different topics; they perform different public functions. Complaint discourse turns technical failure into competence judgment, protest discourse converts visual symbols into solidarity and warning, and electoral discourse makes distrust publicly legible through evaluative monitoring. This matters because computational sentiment analysis often presumes that affect can be stabilized through polarity categories, whereas these findings show that affect is organized by communicative setting. This pattern extends Indonesian sentiment research that has improved classification through transformer and hybrid models [3], [4], [6], but it also exposes a dysfunction: technically correct polarity prediction may still miss what sentiment is doing socially.

This register pattern appears to arise from the structure of Indonesian online discourse itself, where public issues are rarely discussed through neutral information exchange alone. Digital users respond to institutions, elections, scandals, and platform events through evaluative styles that combine complaint, humour, warning, and moral judgment. This does not establish a causal relationship between platform use and affective expression, but it indicates a strong association between topic domain and affective form. Work on Indonesian tweets, YouTube comments, multilingual sentiment datasets, and code-mixing supports this interpretation by showing that domain and language variety shape classification difficulty [5], [17], [10], [8]. The underlying structure is therefore not simply lexical sentiment, but a relation among register, target, stance, and platform convention. This explains why affective labels become unstable when they are detached from discourse context.

The second finding gives sharper significance to irony and sarcasm as mechanisms that reorganize public sentiment. Their function is not ornamental; they allow users to criticize, ridicule, or signal group alignment while avoiding direct propositional accusation. Mock praise, meme-like exaggeration, rhetorical

alarm, and antagonistic joking complicate computational interpretation because they produce a gap between surface wording and intended stance. This finding aligns with sarcasm and irony research showing that pragmatic reversal often defeats models that rely on literal polarity or isolated lexical cues [11], [12], [18]. Yet Indonesian discourse adds a specific challenge: irony is intertwined with slang, collective symbols, code-switching, and locally intelligible references. The implication is that irony auditing should not be treated as a secondary sentiment task. It is central to understanding how public criticism is affectively performed. Fawaid [35] frames language, media, and representation as a politics of who is positioned to speak and be heard, which is consistent with reading ironic criticism as a public performance.

The reason irony becomes so consequential is that Indonesian online interaction often relies on indirectness, shared event knowledge, and culturally recognizable cues. Expressions such as ridicule, exaggerated praise, and symbolic warning work because audiences infer what is meant from context, not because sentiment is transparently encoded in words. This structure creates a correlation between pragmatic density and classification risk. When an item contains mock evaluation or humorous attack, sentiment polarity cannot be read independently from stance target and interactional frame. Cross-linguistic studies of irony detection similarly show that models struggle when irony depends on context outside the sentence or on language-specific conventions [23], [22], [21]. In this corpus, such difficulty is intensified by Indonesian sociolinguistic practices, where affect may be compressed into informal morphology, regionalized phrasing, platform humour, or visual shorthand. Thus, the structure of irony is relational rather than lexical.

The third finding reframes model error as an interpretive phenomenon, not merely a technical failure. Surface polarity bias, register-blind classification, context dependency, multimodal omission, and pragmatic reversal show that misalignment occurs when computational systems detach labels from social meaning. This has important implications for Indonesian NLP evaluation. A model may appear adequate if judged only by accuracy, but its usefulness is weakened if it cannot distinguish complaint from protest, legal explanation from distrust, or humour from hostility. Reviews of sentiment and emotion analysis have repeatedly identified ambiguity, domain shift, bias, and interpretability as persistent challenges [14], [24], [25], [26]. This corpus demonstrates why those concerns are not abstract methodological cautions. In Indonesian online discourse, interpretive error can alter what counts as public emotion, public criticism, or public risk. Abdullah et al. [36] study visual representation in the 2024 Indonesian presidential campaign, a reminder that electoral meaning circulates through images and symbols as well as words.

This misalignment is associated with a deeper structure in current computational modeling: many systems are optimized to predict labels, whereas public discourse requires explanation of why a label is appropriate. Transformer-based and large-language-model approaches have expanded capacity for contextual representation, but they do not automatically solve cultural grounding, multilingual variation, or pragmatic inference. Indonesian NLP studies on low-resource dialects, multilingual reasoning, cultural safety, hallucination, and sociolinguistic bias show that language technologies remain uneven across regional and cultural conditions [7], [30], [31], [34], [9]. Broader computational social science makes a similar point: models can support social analysis only when paired with validation, transparent annotation, and critical interpretation [32], [13]. This study therefore argues that affective modeling in Indonesian online discourse must integrate classification with register analysis, irony auditing, and model-discourse error interpretation.

6. CONCLUSION

This study demonstrates that affect in Indonesian online discourse cannot be adequately understood through polarity classification alone. Its most important lesson is that emotion, irony, and public sentiment operate through registers: civic complaint, protest symbolism, electoral distrust, official/legal explanation, misinformation correction, and fandom conflict each organize affect differently. This contribution refreshes Indonesian computational discourse research by connecting sentiment modeling with register analysis, irony auditing, and model–discourse misalignment. Rather than treating model error as a purely technical problem, this study shows that misclassification may emerge from sociolinguistic density, pragmatic reversal, multimodal symbolism, and context-dependent public meaning.

This study is limited by its reliance on a bounded public corpus and by the absence of live platform interaction data, longitudinal tracking, and large-scale model benchmarking across multiple Indonesian languages. Its findings therefore should not be generalized as a population-level account of Indonesian public opinion. Future research should expand corpus size, compare platforms, incorporate regional languages, test multiple transformer and large-language-model architectures, and include inter-annotator validation for irony, emotion, and stance. Further work should also examine visual and audiovisual affect more systematically, since symbols, screenshots, memes, and video frames often carry evaluative force that text-only sentiment models cannot fully capture.

ACKNOWLEDGMENTS

The author thanks the reviewers and editorial team for their constructive feedback on this manuscript.

FUNDING INFORMATION
Author states no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT
Ivan Riyadi: conceptualization, methodology, formal analysis, writing – original draft, writing – review and editing.

CONFLICT OF INTEREST STATEMENT
Author states no conflict of interest.

AI USE DECLARATION
The author used artificial intelligence tools to assist with the generation of illustrative figures and with the formatting and layout of this manuscript. AI was not used to generate the substantive content, conceptualization, data, analysis, or conclusions of the research, all of which remain the sole responsibility of the author.

INFORMED CONSENT
Not applicable – this study did not involve direct human participants. Data consisted solely of publicly accessible, dateable, and verifiable digital text materials.

ETHICAL APPROVAL
This research did not involve human or animal subjects; it analyzed publicly accessible digital text records and scholarly materials, and therefore no institutional review board approval was required. Where research does involve human subjects, the author affirms that such research would be conducted in full compliance with all relevant national regulations and institutional policies, in accordance with the tenets of the Declaration of Helsinki, and would be approved by the author's institutional review board or equivalent committee.

DATA AVAILABILITY
The corpus of publicly accessible digital text records and scholarly materials analyzed in this study are available from the corresponding author upon reasonable request.

REFERENCES
[1] APJII, *APJII: Jumlah pengguna internet Indonesia tembus 221 juta orang*, Asosiasi Penyelenggara Jasa Internet Indonesia, 2024. [Online]. Available: https://apjii.or.id/berita/d/apjii-jumlah-pengguna-internet-indonesia-tembus-221-juta-orang.
[2] DataReportal, *Digital 2025: Indonesia*, 2025. [Online]. Available: https://datareportal.com/reports/digital-2025-indonesia.
[3] W. Novitasari, "Sentiment analysis on Indonesian political discourse using transformer-based models," *Indonesian Journal of Computational Language Studies*, vol. 1, no. 1, pp. 11-21, 2026.
[4] H. Ahmadian, T. F. Abidin, H. Riza, and K. Muchtar, "Hybrid models for emotion classification and sentiment analysis in Indonesian language," *Applied Computational Intelligence and Soft Computing*, vol. 2024, 2024, doi: 10.1155/2024/2826773.
[5] A. Glenn, P. M. LaCasse, and B. A. Cox, "Emotion classification of Indonesian tweets using bidirectional LSTM," *Neural Computing and Applications*, vol. 35, pp. 9567-9578, 2023, doi: 10.1007/s00521-022-08186-1.
[6] C. Shaw, P. LaCasse, and L. E. Champagne, "Exploring emotion classification of Indonesian tweets using large scale transfer learning via IndoBERT," *Social Network Analysis and Mining*, vol. 15, 2025, doi: 10.1007/s13278-025-01439-6.
[7] M. I. Ibrahim, "Morphological segmentation of low-resource Indonesian dialects using unsupervised neural models," *Indonesian Journal of Computational Language Studies*, vol. 1, no. 1, pp. 53-62, 2026.
[8] A. Nugroho, "Code-mixing detection in Indonesian-English tweets using machine learning and linguistic features," *Indonesian Journal of Computational Language Studies*, vol. 1, no. 1, pp. 22-31, 2026.
[9] M. A. Sumadi, "Who gets represented by Indonesian AI? Measuring regional, gender, and sociolinguistic bias in large language models," *Indonesian Journal of Computational Language Studies*, vol. 1, no. 2, 2026.
[10] G. I. Winata, A. F. Aji, S. Cahyawijaya, R. Mahendra, F. Koto, A. Romadhony, K. Kurniawan, D. Moeljadi, R. E. Prasojo, P. Fung, T. Baldwin, J. H. Lau, R. Sennrich, and S. Ruder, "NusaX: Multilingual parallel sentiment dataset for 10 Indonesian local languages," 2022, doi: 10.48550/arXiv.2205.15960.
[11] R. Akula and I. I. Garibay, "Interpretable multi-head self-attention architecture for sarcasm detection in social media," *Entropy*, vol. 23, no. 4, 2021, doi: 10.3390/e23040394.
[12] S. Frenda, A. T. Cignarella, V. Basile, C. Bosco, V. Patti, and P. Rosso, "The unbearable hurtfulness of sarcasm," *Expert Systems with Applications*, vol. 193, p. 116398, 2022, doi: 10.1016/j.eswa.2021.116398.
[13] C. Ziems, W. B. Held, O. Shaikh, J. Chen, Z. Zhang, and D. Yang, "Can large language models transform computational social science?," *Computational Linguistics*, vol. 50, pp. 237-291, 2023, doi: 10.1162/coli_a_00502.
[14] P. Nandwani and R. Verma, "A review on sentiment analysis and emotion detection from text," *Social Network Analysis and Mining*, vol. 11, 2021, doi: 10.1007/s13278-021-00776-6.
[15] S. Kusal, S. Patil, K. Kotecha, R. Aluvalu, and V. Vijayakumar, "AI based emotion detection for textual big data: Techniques and contribution," *Big Data and Cognitive Computing*, vol. 5, no. 3, p. 43, 2021, doi: 10.3390/bdcc5030043.

[16] K. Machová, M. Szabóová, J. Paralič, and J. Mičko, "Detection of emotion by text analysis using machine learning," *Frontiers in Psychology*, vol. 14, 2023, doi: 10.3389/fpsyg.2023.1190326.
[17] M. Hidayat, D. D. Prasetya, and T. Widiyaningtyas, "Sentiment analysis of Latinized Arabic and emoji in Indonesian YouTube comments: A LABERT-LSTM model," *Journal of Applied Engineering and Technological Science*, 2025, doi: 10.37385/jaets.v6i2.7000.
[18] D. Olaniyan, R. Ogundokun, O. P. Bernard, J. Olaniyan, R. Maskeliūnas, and H. B. Akande, "Utilizing an attention-based LSTM model for detecting sarcasm and irony in social media," *Computers*, vol. 12, no. 11, p. 231, 2023, doi: 10.3390/computers12110231.
[19] Y. Y. Tan, C. Chow, J. Kanesan, J. H. Chuah, and Y. Lim, "Sentiment analysis and sarcasm detection using deep multi-task learning," *Wireless Personal Communications*, vol. 129, pp. 2213-2237, 2023, doi: 10.1007/s11277-023-10235-4.
[20] A. D. Yacoub, A. Aboutabl, and S. O. Slim, "Multilingual sarcasm detection for enhancing sentiment analysis using deep learning algorithms," *Journal of Communications Software and Systems*, 2024, doi: 10.24138/jcomss-2024-0071.
[21] P. Yi, Y. Xia, and Y. Long, "Irony detection, reasoning, and understanding in zero-shot learning," *IEEE Transactions on Artificial Intelligence*, vol. 7, pp. 740-754, 2025, doi: 10.1109/tai.2025.3579452.
[22] N. Bölücü and B. Can, "Semantically-informed graph neural networks for irony detection in Turkish," *ACM Transactions on Asian and Low-Resource Language Information Processing*, vol. 24, pp. 1-20, 2024, doi: 10.1145/3705610.
[23] A. Maladry, E. Lefever, C. Van Hee, and V. Hoste, "The limitations of irony detection in Dutch social media," *Language Resources and Evaluation*, pp. 1-32, 2023, doi: 10.1007/s10579-023-09656-1.
[24] M. Wankhade, A. C. Rao, and C. Kulkarni, "A survey on sentiment analysis methods, applications, and challenges," *Artificial Intelligence Review*, vol. 55, pp. 5731-5780, 2022, doi: 10.1007/s10462-022-10144-1.
[25] M. Kumar, L. Khan, and H.-T. Chang, "Evolving techniques in sentiment analysis: A comprehensive review," *PeerJ Computer Science*, vol. 11, 2025, doi: 10.7717/peerj-cs.2592.
[26] X. Chen, H. Xie, X. Tao, F. Wang, D. Zhang, and H. Dai, "A computational analysis of aspect-based sentiment analysis research through bibliometric mapping and topic modeling," *Journal of Big Data*, vol. 12, 2025, doi: 10.1186/s40537-025-01068-y.
[27] J. Santoso, "Indonesian named entity recognition using BiLSTM-CRF with domain-specific word embeddings," *Indonesian Journal of Computational Language Studies*, vol. 1, no. 1, pp. 1-10, 2026.
[28] R. Perwira, V. A. Permadi, D. I. Purnamasari, and R. P. Agusdin, "Domain-specific fine-tuning of IndoBERT for aspect-based sentiment analysis in Indonesian travel user-generated content," *Journal of Information Systems Engineering and Business Intelligence*, 2025, doi: 10.20473/jisebi.11.1.30-40.
[29] A. Jazuli and R. Kusumaningrum, "Optimizing aspect-based sentiment analysis using BERT for comprehensive analysis of Indonesian student feedback," *Applied Sciences*, 2024, doi: 10.3390/app15010172.
[30] F. D. Wulandari, "Do large language models reason consistently across Indonesian languages? A multilingual evaluation of Indonesian, Javanese, Sundanese, and Buginese," *Indonesian Journal of Computational Language Studies*, vol. 1, no. 2, 2026.
[31] Nadiyah, "When safety fails in local languages: Evaluating culturally sensitive responses of large language models in multilingual Indonesia," *Indonesian Journal of Computational Language Studies*, vol. 1, no. 2, 2026.
[32] D. Demszky, D. Yang, D. S. Yeager, C. J. Bryan, M. Clapper, S. Chandhok, J. Eichstaedt, C. A. Hecht, J. P. Jamieson, M. Johnson, D. Jones Krettek-Cobb, L. Y. Lai, N. JonesMitchell, D. C. Ong, C. Dweck, J. J. Gross, and J. W. Pennebaker, "Using large language models in psychology," *Nature Reviews Psychology*, vol. 2, pp. 688-701, 2023, doi: 10.1038/s44159-023-00241-5.
[33] D. W. Heaton, J. Clos, E. Nichele, and J. Fischer, "Critical reflections on three popular computational linguistic approaches to examine Twitter discourses," *PeerJ Computer Science*, vol. 9, 2023, doi: 10.7717/peerj-cs.1211.
[34] R. T. Wibisono, "Hallucinating the archipelago: Detecting factual and cultural errors in retrieval-augmented generation for Indonesian knowledge," *Indonesian Journal of Computational Language Studies*, vol. 1, no. 2, 2026.
[35] A. Fawaid, *Who Controls the Narrative? Language, Media, and Politics of Representation*. Probolinggo, Indonesia: Pustaka Nurja, 2026.
[36] I. Abdullah, D. Afriadi, M. Yusuf, M. Susanto, and A. Fawaid, "The optics of leadership: Visual representation in the 2024 Indonesian presidential campaign," *Visual Review. International Visual Culture Review*, vol. 17, no. 2, pp. 133-145, 2025, doi: 10.62161/revvisual.v17.5273.

BIOGRAPHY OF AUTHOR



Ivan Riyadi is affiliated with Universitas Islam Negeri Raden Fatah Palembang, Indonesia, where his work engages computational discourse analysis, affective computing, and the pragmatics of irony and sentiment in Indonesian online communication. His current research examines how register-sensitive affect coding, irony auditing, and model–discourse misalignment analysis can improve the cultural validity of computational sentiment classification. He can be contacted at email: ivanriyadi_uin@radenfatah.ac.id