Vol. 11, No. 2, Oktober 2026, Page. 539-553
https://journal.uin-alauddin.ac.id/index.php/instek
p-ISSN: 2541-1179
e-ISSN: 2581-1711
JURNAL INSTEK
INFORMATIKA SAINS DAN TEKNOLOGI

Ekstraksi Relasi Antar Karakter Pada Teks *Satua Bali* Menggunakan *Support Vector Machine* Dengan Pendekatan Fitur Leksikon

I Kadek Dwi Adnyana[*1], Ngurah Agus Sanjaya ER[2], I Ketut Gede Suhartana[3]

[1,2,3]Magister Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Udayana, Indonesia
Email: [1]dekguss08@gmail.com, [2]agus_sanjaya@unud.ac.id, [3]ikg.suhartana@unud.ac.id

Abstrak

Pemetaan relasi sosial antar tokoh pada *Satua Bali* penting untuk mendukung preservasi konteks budaya dan pengembangan *knowledge graph* sastra Nusantara. Penelitian ini mengklasifikasikan relasi Keluarga, Teman, dan Musuh menggunakan *Support Vector Machine* (SVM) dengan arsitektur *Dual-Stream Pipeline* yang menggabungkan TF-IDF *N-gram* pada teks dengan entitas tersamarkan dan fitur leksikon pada teks asli. Teks mentah dikumpulkan dari situs web publik yang memuat cerita *Satua Bali*, sedangkan kontribusi data penelitian terletak pada penyusunan skema dan anotasi manual tiga kelas relasi. Dari 637 segmen teranotasi, deduplikasi menghasilkan 600 segmen unik yang dievaluasi menggunakan *Stratified 5-Fold Cross Validation*. Model *Dual-Stream* mencapai akurasi 75,0% dan *Macro F1* 75,2%, lebih tinggi dibandingkan *baseline* TF-IDF pada teks dengan entitas tersamarkan dengan akurasi dan *Macro F1* sebesar 71,5%. Sementara itu, TF-IDF pada teks asli mencapai akurasi 83,7%, menunjukkan bahwa nama tokoh dan leksikon asli sangat prediktif, tetapi berpotensi membawa bias identitas. Hasil ini menunjukkan bahwa *Dual-Stream* mampu memulihkan sebagian informasi relasional akibat *masking* dan berfungsi sebagai strategi mitigasi terhadap *trade-off entity masking*.

Kata kunci: ekstraksi relasi, *natural language processing*, satua bali, *support vector machine*, TF-IDF

Abstract

Mapping social relations among characters in Satua Bali is important for preserving cultural context and supporting the development of a Nusantara literature knowledge graph. This study classifies Family, Friend, and Enemy relations using a Support Vector Machine (SVM) with a Dual-Stream Pipeline combining TF-IDF N-gram features from masked text and lexicon-based features from the original text. Raw narratives were collected from publicly accessible websites containing Satua Bali stories, while the data contribution lies in the manually constructed three-class relation annotation scheme. From 637 annotated segments, deduplication produced 600 unique segments evaluated using Stratified 5-Fold Cross Validation. The Dual-Stream model achieved 75.0% accuracy and a 75.2% Macro F1, outperforming the masked-text TF-IDF baseline, which obtained 71.5% accuracy and Macro F1. In contrast, TF-IDF on the original text reached 83.7% accuracy, indicating that character names and original lexical cues are highly predictive but may introduce identity bias. These results indicate that the Dual-Stream approach recovers part of the relational information lost through masking and should be interpreted as a mitigation strategy for the entity-masking trade-off rather than a definitive solution.

Keywords: *relation extraction, natural language processing, satua bali, support vector machine, TF-IDF*

This work is an open access article and licensed under a Creative Commons Attribution-NonCommercial ShareAlike 4.0 International (CC BY-NC-SA 4.0)

1. PENDAHULUAN

Satua Bali adalah karya sastra lisan warisan budaya Nusantara yang kaya akan nilai moral dan kearifan lokal, serta secara mendalam mencerminkan pola interaksi sosial masyarakat Bali [1]. Seiring dengan pesatnya digitalisasi di berbagai sektor, pelestarian sastra tradisional kini tidak lagi sebatas pada pendokumentasian teks fisik menjadi format digital, melainkan telah meluas menuju pemahaman linguistik komputasional melalui teknologi *Natural Language Processing* (NLP) [2]. Melalui NLP, analisis semantik terhadap teks sastra dapat dilakukan secara komputasional, sehingga mesin dapat dilatih untuk membaca, memahami, dan mengekstraksi informasi berharga secara otomatis, termasuk memetakan jejaring relasi sosial antar karakter di dalam cerita rakyat. Pemetaan relasi ini menjadi sangat mendesak untuk dilakukan secara otomatis guna mencegah erosi pemahaman konteks budaya di era

modern, sekaligus bertindak sebagai fondasi krusial dalam pembangunan graf pengetahuan (*knowledge graph*) sastra Nusantara di masa depan.

Meskipun teknologi ekstraksi informasi telah berkembang pesat, penerapannya pada bahasa daerah seperti bahasa Bali masih menghadapi hambatan yang sangat krusial. Masalah utama dalam pemrosesan bahasa daerah adalah statusnya sebagai *low-resource language* (bahasa dengan sumber daya terbatas), di mana korpus beranotasi masif sangat langka [3]. Selain keterbatasan data, kompleksitas morfologi bahasa Bali menambah tantangan komputasi. Penggunaan afiksasi atau imbuhan penunjuk kepemilikan (misalnya akhiran *-ne* pada kata "*memene*" atau "*pianakne*") lazim digunakan dalam *Satua Bali* untuk menunjukkan status interaksi sosial. Kesalahan algoritma dalam mengenali pola linguistik spesifik ini berpotensi menyebabkan mesin gagal membedakan apakah tokoh-tokoh tersebut memiliki relasi keluarga, pertemanan, atau permusuhan.

Pendekatan *machine learning* sering digunakan dalam klasifikasi teks, di mana tahapan utamanya tidak terlepas dari proses ekstraksi fitur. Salah satu algoritma klasifikasi yang dikenal paling andal dalam memproses data teks adalah *Support Vector Machine* (SVM). SVM dikenal memiliki performa yang baik pada data teks karena kemampuannya dalam menangani data yang kompleks. SVM bekerja dengan mencari *hyperplane* dengan margin terbesar untuk memisahkan antar kelas secara optimal. Untuk merepresentasikan teks menjadi vektor numerik, *Term Frequency-Inverse Document Frequency* (TF-IDF) merupakan metode ekstraksi fitur yang efektif karena melakukan pembobotan kata berdasarkan frekuensi kemunculan serta kelangkaannya. Namun, metode TF-IDF konvensional rentan kehilangan pemahaman konteks semantik spesifik kata. Oleh karena itu, pendekatan kombinasi (*hybrid*) yang mengintegrasikan fitur leksikon kustom (*Lexicon-based feature*) diperlukan untuk menyuntikkan probabilitas kemunculan kata kunci penentu relasi secara eksplisit ke dalam algoritma [4].

Tinjauan terhadap penelitian sebelumnya memperlihatkan dominasi eksplorasi NLP pada teks berbahasa Indonesia arus utama, namun menyajikan analisis komparatif algoritma yang relevan. Penelitian oleh Noor dan Normawati membandingkan kinerja SVM dan *K-Nearest Neighbor* (KNN) menggunakan ekstraksi fitur TF-IDF, *Bag-of-Words* (BoW), dan Word2Vec [5]. Hasil pengujian menunjukkan bahwa SVM lebih unggul daripada KNN dalam analisis sentimen. Pada kombinasi SVM dan TF-IDF, akurasi mencapai 81%, yang membuktikan bahwa pendekatan SVM dengan kernel dan optimasi margin terbukti lebih stabil terhadap keberagaman data (*noise*) dibandingkan KNN.

Keunggulan metode TF-IDF yang digabungkan dengan algoritma SVM juga dikonfirmasi oleh riset terbaru Difandana dkk. (2026) dan Fajriah dkk. (2026), di mana model SVM berhasil mencapai presisi di atas 92% dalam mendeteksi *hate speech* dan pesan teks berbahaya berdimensi tinggi [6], [7]. Di sisi lain, untuk mengatasi kelemahan model dalam menangkap semantik linguistik, penelitian oleh Nurcahyawati & Mustaffa (2023) serta Muhammadi dkk. (2022) berhasil membuktikan bahwa integrasi injeksi aturan berbasis kamus (*Lexicon-based resources*) ke dalam SVM mampu meningkatkan akurasi klasifikasi secara signifikan [8], [9]. Akan tetapi, dalam ranah yang lebih spesifik seperti *Relation Extraction* pada bahasa sumber daya rendah, studi eksperimental oleh Wan dkk. (2025) menunjukkan bahwa pendekatan model berbasis *deep learning* pun masih kesulitan menalar relasi tanpa bantuan anotasi logis dari aturan berbasis pengetahuan (*knowledge-based rules*) yang spesifik [10]. Perkembangan representasi kontekstual juga menyediakan arah pembanding yang lebih mutakhir. IndoBERT dikembangkan sebagai model bahasa pralatih untuk berbagai tugas NLP berbahasa Indonesia [11], sedangkan NusaBERT memperluas IndoBERT melalui perluasan kosakata dan pralatih pada korpus multibahasa yang mencakup bahasa-bahasa daerah Indonesia [12]. Meskipun kedua model tersebut belum diuji pada eksperimen ini, pendekatan tersebut menjadi pembanding *state-of-the-art* yang relevan untuk pengembangan ekstraksi relasi *Satua Bali* selanjutnya.

Berdasarkan tinjauan literatur di atas, terdapat celah penelitian (*research gap*), yakni ekstraksi relasi pada teks sastra kedaerahan bersumber daya rendah seperti *Satua Bali* masih minim dieksplorasi [13]. Penggunaan *entity masking* juga menimbulkan *trade-off*: anonimisasi nama dapat mengurangi ketergantungan model terhadap identitas tokoh, tetapi dapat menurunkan informasi entitas yang relevan bagi klasifikasi relasi [14]. Karena itu, dampak *masking* pada penelitian ini diuji secara empiris melalui pembandingan representasi teks dengan entitas tersamarkan dan teks asli.

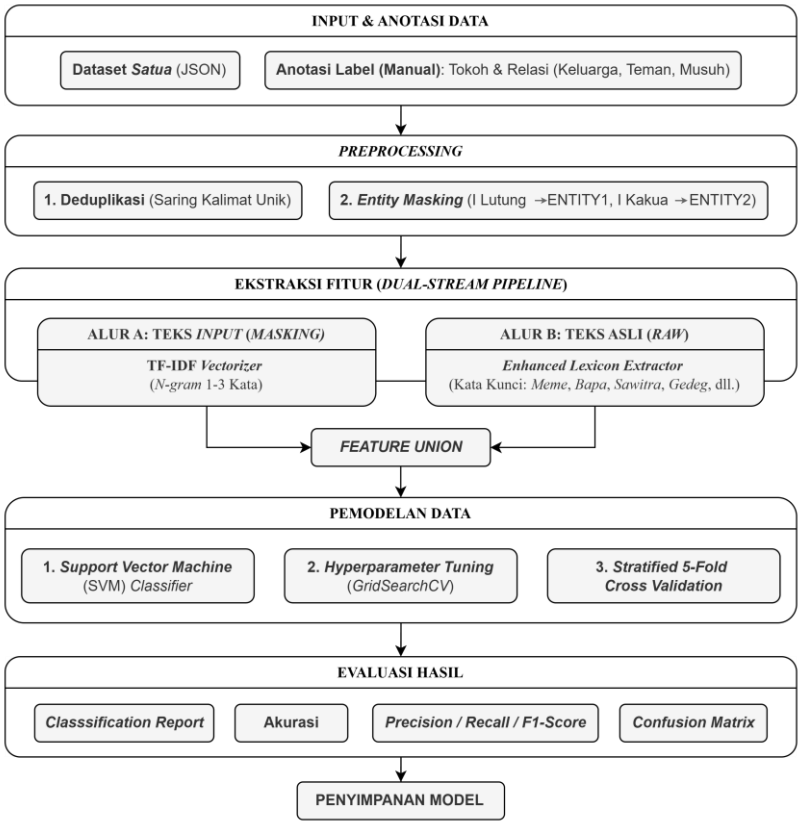
Kebaruan penelitian ini terletak pada dua aspek. Pertama, penelitian membentuk lapisan anotasi baru berupa pasangan tokoh dan tiga kelas relasi, yaitu Keluarga, Teman, dan Musuh, pada teks mentah *Satua Bali* yang berasal dari sumber web publik. Kedua, penelitian mengusulkan arsitektur *Dual-Stream*

yang menggabungkan TF-IDF pada teks dengan entitas tersamarkan dengan fitur leksikon dari teks asli untuk mengurangi kehilangan informasi relasional akibat *masking*.

Tujuan utama penelitian ini adalah mengklasifikasikan relasi antar tokoh (Keluarga, Teman, dan Musuh) dalam teks *Satua Bali* menggunakan algoritma *Support Vector Machine*. Kontribusi penelitian mencakup penyusunan skema anotasi relasi, perancangan *Dual-Stream Pipeline*, serta evaluasi empiris untuk mengukur kontribusi fitur leksikon dan konsekuensi penggunaan *masking*.

2. METODE PENELITIAN

Penelitian ini menggunakan pendekatan *supervised machine learning* untuk melakukan klasifikasi relasi antar tokoh pada teks *Satua Bali*. Proses klasifikasi dirancang menggunakan arsitektur *Dual-Stream Pipeline* yang menggabungkan fitur statistik dan fitur berbasis leksikon sebelum diumpankan ke dalam model *Support Vector Machine* (SVM). Alur penelitian diilustrasikan pada Gambar 1.



Gambar 1. Alur Penelitian

2.1. Pengumpulan dan Anotasi Data

Data yang digunakan dalam penelitian ini berupa teks mentah dari 30 cerita *Satua Bali* yang dikumpulkan dari situs web publik yang memuat cerita rakyat Bali. Teks mentah tersebut bukan merupakan kontribusi data baru penelitian. Kontribusi dataset pada penelitian ini terletak pada proses kurasi dan pembentukan lapisan anotasi relasi antar tokoh yang sebelumnya tidak tersedia dalam bentuk tiga kelas Keluarga, Teman, dan Musuh. Pengumpulan data difokuskan pada segmen naratif yang memuat interaksi, komunikasi, hubungan kekerabatan, kerja sama, atau konflik antara dua tokoh sehingga dapat dibentuk sebagai unit klasifikasi terawasi [15].

Setelah teks mentah (*raw text*) terkumpul, anotasi dilakukan secara manual dengan menetapkan pasangan Tokoh 1 dan Tokoh 2 serta satu label relasi yang paling sesuai dengan bukti relasional pada segmen. Seluruh kalimat atau segmen pendek yang memenuhi kriteria relasional diambil; apabila pasangan tokoh yang sama muncul pada beberapa bagian cerita, setiap segmen tetap dipertahankan sebagai sampel terpisah selama memuat bukti relasional yang dapat diidentifikasi. Pada segmen yang menyebut lebih dari dua tokoh, anotasi difokuskan pada pasangan yang hubungan relasionalnya

dinyatakan paling jelas. Pedoman anotasi menggunakan definisi operasional tiga kelas sebagaimana dijelaskan pada bagian berikut dan dicontohkan pada Tabel 1. Data hasil anotasi kemudian distrukturkan dalam format *JavaScript Object Notation* (JSON) untuk memudahkan pemrosesan komputasional.

Unit analisis dibatasi pada satu kalimat atau segmen naratif pendek yang berpusat pada satu pasangan tokoh. Batas ini digunakan untuk menjaga konsistensi anotasi serta memungkinkan pembandingan fitur TF-IDF dan leksikon secara terkontrol. Dengan demikian, label merepresentasikan bukti relasi pada konteks lokal, bukan keseluruhan keadaan hubungan tokoh sepanjang alur cerita. Setiap sampel data di dalam berkas JSON tersebut mendefinisikan empat pasangan *key-value* utama, yaitu:

- Teks, yaitu kalimat utuh asli yang disalin dari sumber cerita Satua Bali.
- Tokoh 1, yaitu entitas karakter pertama yang menjadi subjek pelibat dalam interaksi.
- Tokoh 2, yaitu entitas karakter kedua yang menjadi objek atau subjek interaksi.
- Relasi, yaitu label kelas target yang digunakan sebagai referensi supervisi dan merepresentasikan hubungan sosial antara Tokoh 1 dan Tokoh 2 pada unit teks yang dianotasi.

Label relasi dikategorikan ke dalam tiga kelas utama yang mencerminkan pola interaksi sosial dominan dalam budaya dan narasi *Satua Bali*, yaitu:

- Keluarga: Merupakan label untuk interaksi yang melibatkan ikatan darah, garis keturunan, atau ikatan perkawinan. Ini mencakup hubungan seperti orang tua dengan anak (*meme/bapa* dan *pianak*), kakek/nenek dengan cucu (*dadong/pekak* dan *cucu*), hingga suami istri (*kurenan/somah*).
- Teman: Merupakan label untuk interaksi yang menunjukkan persahabatan (*masawitra*), kerja sama, saling menolong, atau bahkan hubungan hierarkis yang positif seperti antara raja dan pelayannya (*anake agung* dan *parekan*).
- Musuh: Merupakan label untuk interaksi yang memuat konflik, kemarahan (*brangti/gedeg*), niat buruk, penipuan (*nguluk-nguluk*), hingga pertarungan fisik atau pembunuhan (*matiang*).

Validasi anotasi pada penelitian ini dilakukan sebagai pemeriksaan konsistensi internal. Pemeriksaan mencakup kelengkapan pasangan entitas, kesesuaian nama tokoh dengan teks, konsistensi label terhadap definisi operasional, peninjauan ulang terhadap segmen yang memuat indikator lintas kelas, serta penghapusan duplikasi sebelum pemodelan. Setelah deduplikasi, diperoleh 600 segmen unik dengan distribusi seimbang, yaitu masing-masing 200 segmen untuk kelas Keluarga, Teman, dan Musuh. Karena proses anotasi yang terdokumentasi pada penelitian ini belum melibatkan anotator independen kedua, nilai *inter-annotator agreement* belum dilaporkan; kondisi tersebut diperlakukan sebagai keterbatasan validasi data dan tidak digunakan sebagai dasar untuk mengklaim reliabilitas antar-anotator.

Tabel 1. Anotasi Data

Teks	Tokoh 1	Tokoh 2	Relasi
Ada tuturan satua I Rajapala makurenan ngajak dedari madan Ken Sulasih tur ngelah pianak adiri.	*I Rajapala*	*Ken Sulasih*	Keluarga
I Lutung masawitra ajak I Kakua tur bareng-bareng mamaling isen di tegal dukuh Kantrungan.	*I Lutung*	*I Kakua*	Teman
Sedeng iteha I Lutung ngamah biu lantas katumbak baan I Kaki Perodong beneng lambungne kanti ulung.	*I Lutung*	*I Kaki Perodong*	Musuh

Vol. 11, No. 2, Oktober 2026, Page. 539-553
https://journal.uin-alauddin.ac.id/index.php/instek
p-ISSN: 2541-1179
e-ISSN: 2581-1711
JURNAL INSTEK
INFORMATIKA SAINS DAN TEKNOLOGI

2.2. Pra-pemrosesan Data (*Data Preprocessing*)

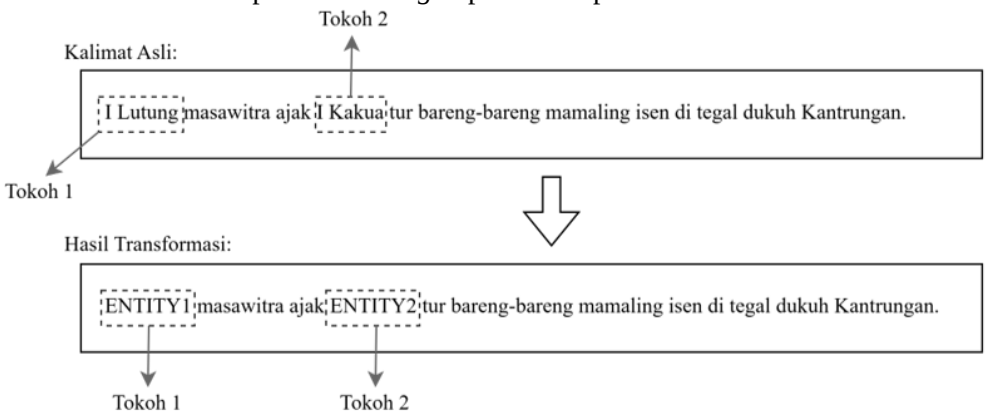
Pra-pemrosesan bertujuan untuk mereduksi *noise* dan menyiapkan teks ke dalam format yang dapat dipahami secara optimal oleh algoritma klasifikasi. Berbeda dengan pemrosesan teks standar yang sering kali membuang banyak struktur kalimat, penelitian ini mempertahankan keutuhan semantik melalui dua teknik utama yang berfokus pada kualitas dan integritas konteks:

a. Deduplikasi Data (*Data Deduplication*)

Deduplikasi merupakan proses pembersihan tingkat korpus yang dilakukan melalui penyaringan struktur kamus data (*dictionary*). Sistem memindai seluruh atribut teks di dalam format JSON, dan jika ditemukan lebih dari satu entri dengan struktur kalimat yang sama persis, sistem hanya akan mempertahankan satu instans unik serta menghapus salinannya. Langkah ini krusial untuk mencegah kebocoran data (*data leakage*) antara subset pelatihan dan pengujian, yang sering menjadi penyebab utama terjadinya *overfitting* [16]. Pada implementasi penelitian ini, proses deduplikasi berhasil menyaring kumpulan data mentah yang awalnya berjumlah 637 kalimat teranotasi, menjadi 600 kalimat unik yang siap diproses ke tahap selanjutnya.

b. *Entity Masking*

Entity masking adalah teknik anonimisasi berbasis aturan (*rule-based replacement*) yang diterapkan secara spesifik pada kata benda diri (*proper noun*) karakter. Fungsi komputasional akan membaca nilai teks dari atribut Tokoh 1 dan Tokoh 2, kemudian melacak keberadaan string tersebut secara presisi di dalam kalimat, untuk selanjutnya ditimpa dengan token statis berformat ENTITY1 dan ENTITY2 [17], [18]. Tujuan utama dari teknik ini adalah untuk menghindari bias algoritmik. Tanpa *masking*, model *machine learning* cenderung sekadar menghafal asosiasi nama tokoh alih-alih mempelajari polanya (contoh: model bias menganggap "*I Lutung*" pasti berlabel musuh karena sering muncul di kalimat bermuatan konflik). Dengan *masking*, model dipaksa untuk fokus mengekstraksi pemahaman dari kata kerja, kata sifat, maupun struktur sintaksis yang mengelilingi entitas tersebut. Contoh hasil transformasi dari proses *masking* dapat dilihat pada Gambar 2.



Gambar 2. Proses *Masking*

2.3. Ekstraksi Fitur Berbasis *Dual-Stream*

Entity *masking* digunakan untuk mengurangi ketergantungan model terhadap identitas tokoh, tetapi teknik ini juga dapat mengurangi informasi entitas yang relevan bagi ekstraksi relasi [14]. Oleh karena itu, kondisi tersebut diperlakukan sebagai *trade-off* yang harus diuji secara empiris, bukan sebagai permasalahan yang diasumsikan selesai melalui perubahan arsitektur.

Untuk memitigasi degradasi tersebut, penelitian ini menggunakan *Dual-Stream Pipeline* melalui modul *FeatureUnion*. Aliran pertama memproses TF-IDF pada teks dengan entitas tersamarkan untuk menangkap pola konteks di sekitar entitas, sedangkan aliran kedua mengekstrak fitur leksikon dari teks asli. Kontribusi aliran leksikon kemudian diukur terhadap *baseline* tanpa leksikon.

a. Aliran Pertama: *Term Frequency-Inverse Document Frequency* (TF-IDF)

Aliran pertama secara khusus memproses teks dari variabel $teks_input$ (teks yang telah di-*masking*). Ekstraksi ini menggunakan *N-Gram* TF-IDF dengan rentang (1, 3), yang berarti sistem akan mengekstrak kombinasi satu kata (*unigram*), dua kata (*bigram*), hingga tiga kata berurutan (*trigram*) [19]. Kombinasi *N-gram* ini memungkinkan model menangkap frasa kontekstual di sekitar token *masking*, seperti "ENTITY1 *ngajakin* ENTITY2" [20], [21]. TF-IDF mengukur seberapa penting sebuah kata terhadap suatu dokumen dalam kumpulan korpus. Bobot TF-IDF dari sebuah term t dalam dokumen d dirumuskan seperti Persamaan (1).

$$W_{t,d} = tf_{t,d} \times log\left(\frac{N}{df_t}\right) \quad (1)$$

Dalam Persamaan (1), variabel $W_{t,d}$ merepresentasikan bobot kalkulasi akhir dari term t pada dokumen d. Nilai ini dipengaruhi oleh $tf_{t,d}$, yakni frekuensi kemunculan term t di dalam dokumen d, yang secara spesifik merepresentasikan seberapa dominan sebuah kata dalam satu kalimat tunggal. Selain itu, formula ini mengikutsertakan perbandingan antara total jumlah kalimat di dalam korpus (N) dengan jumlah dokumen yang mengandung term t (df_t). Penggunaan fungsi logaritma $\log(N/df_t)$ pada operan tersebut secara matematis bertindak sebagai penalti yang akan memberikan bobot rendah pada kata-kata sambung umum yang sering muncul melintasi berbagai kalimat. Sebaliknya, mekanisme ini akan memberikan amplifikasi bobot yang tinggi pada kata-kata spesifik yang sangat diskriminatif dan krusial dalam membedakan antar kelas relasi.

b. Aliran Kedua: *Enhanced Lexicon Extractor*

Sebagai pelengkap dari aliran pertama, aliran kedua bekerja murni mengekstrak data dari variabel $teks_asli$ (teks mentah tanpa *masking*). Aliran ini menggunakan kelas ekstraktor kustom (*rule-based*) bernama *EnhancedLexiconExtractor* [22], [23]. Algoritma ini dirancang khusus untuk memindai kalimat dan menghitung frekuensi kemunculan lema yang merujuk pada tiga relasi kultural Bali: Keluarga (misal: *meme, bapa, pianak, kurenan*), Teman (misal: *masawitra, parekan, braya*), dan Musuh (misal: *gedeg, matiang, puun*). Skor leksikon untuk kelas C dihitung menggunakan Persamaan (2).

$$Score_C(d) = \sum_{i=1}^{k} \mathbb{I}\,(w_i \in L_C) \times \beta \quad (2)$$

Dalam Persamaan (2), proses pencocokan menggunakan fungsi indikator yang bernilai 1 ketika kata ditemukan pada himpunan leksikon kelas relasi dan 0 ketika tidak ditemukan. Akumulasi indikator kemudian dikalikan dengan bobot penguat B untuk mengatur kontribusi relatif fitur leksikon terhadap fitur TF-IDF [24]. Nilai B diuji pada himpunan {1,0; 1,5; 2,0; 2,5; 3,0} menggunakan skema validasi yang sama. Berdasarkan hasil uji sensitivitas, nilai B = 2,0 dipilih karena berada pada rentang akurasi yang stabil tanpa menimbulkan penguatan leksikon secara berlebihan. Selanjutnya, *FeatureUnion* menggabungkan matriks TF-IDF dan tiga skor leksikon sebelum representasi tersebut diumpankan ke model *Support Vector Machine* (SVM).

2.4. Metode Pembanding dan Uji Ablasi

Untuk mengukur kontribusi aliran leksikon dan besarnya perubahan performa akibat masking, model usulan dibandingkan dengan dua *baseline* yang menggunakan *classifier* dan prosedur evaluasi yang sama.

a. TF-IDF Teks dengan Entitas Tersamarkan + SVM

Baseline pertama hanya menggunakan teks dengan entitas yang telah disamarkan melalui *entity masking* dan direpresentasikan dengan TF-IDF *N-gram* (1, 3), tanpa fitur leksikon. Skenario ini berfungsi sebagai uji ablasi untuk mengukur performa ketika aliran leksikon pada *Dual-Stream* dihilangkan.

b. TF-IDF Teks Asli + SVM

Baseline kedua menggunakan teks asli tanpa masking dan direpresentasikan dengan TF-IDF *N-gram* (1, 3), tanpa fitur leksikon terpisah. Skenario ini digunakan untuk mengukur seberapa besar informasi prediktif yang tersedia ketika nama tokoh dan seluruh leksikon asli tetap dipertahankan.

Seluruh skenario menggunakan ruang pencarian hiperparameter SVM yang sama, yaitu kernel linear dan RBF, nilai C sebesar 0,1; 1; 10; dan 50, serta nilai gamma berupa scale; 0,1; dan 1. Evaluasi juga menggunakan *Stratified 5-Fold Cross Validation* dengan pengacakan yang sama (random state 42), sehingga perbedaan kinerja terutama merefleksikan perbedaan representasi masukan dan keberadaan aliran leksikon

2.5. Klasifikasi *Support Vector Machine (SVM)*

Fitur gabungan yang telah diekstraksi dari TF-IDF dan leksikon kemudian diumpankan ke dalam algoritma *Support Vector Machine* (SVM). SVM merupakan algoritma *machine learning* yang mengklasifikasikan data dengan cara mencari *hyperplane* (bidang pemisah) optimal yang menghasilkan margin maksimum antar kelas [4], [25], [26]. Memaksimalkan margin ini sangat penting karena berfungsi untuk meminimalkan *generalization error*, sehingga model tetap tangguh dan akurat ketika dihadapkan pada data baru yang belum pernah dipelajari sebelumnya.

Mengingat penelitian ini merupakan masalah *multiclass* dengan 3 kelas target (Keluarga, Teman, Musuh), algoritma menerapkan pendekatan *One-vs-One* (OvO). Melalui pendekatan ini, sistem tidak memisahkan ketiga kelas sekaligus, melainkan membangun $\frac{N(N-1)}{2}$ atau 3 pengklasifikasi biner terpisah (Keluarga vs Teman, Keluarga vs Musuh, dan Teman vs Musuh). Hasil prediksi akhir kemudian ditarik berdasarkan mekanisme *majority voting* (suara terbanyak) dari ketiga model biner tersebut. Secara matematis, *hyperplane* linear untuk memisahkan data dirumuskan seperti Persamaan (3).

$$f(x) = w^T x + b = 0 \qquad (3)$$

Dalam formulasi Persamaan (3), variabel *w* merepresentasikan vektor bobot (*weight vector*) yang posisinya tegak lurus terhadap *hyperplane* dan berfungsi secara krusial untuk menentukan orientasi dari bidang pemisah antar kelas. Vektor bobot ini dikalikan secara *dot product* dengan variabel x, yang merupakan vektor fitur dari representasi teks masukan. Selain itu, persamaan ini mengikutsertakan variabel *b* yang merepresentasikan nilai bias, di mana nilai ini berperan dalam menentukan seberapa jauh *hyperplane* tersebut mengalami pergeseran dari titik asal (*origin*) di dalam ruang dimensi fitur.

Karakteristik data teks sering kali memiliki distribusi yang sangat kompleks dan tidak selalu dapat dipisahkan secara linear di ruang fitur aslinya. Untuk mengatasi limitasi tersebut, SVM menerapkan *kernel trick*, yaitu sebuah fungsi matematika yang mentransformasi data ke dalam ruang dimensi yang lebih tinggi agar batas keputusan non-linear dapat ditarik. Salah satu fungsi kernel yang dieksplorasi dalam penelitian ini adalah *Radial Basis Function* (RBF) [27], dengan Persamaan (4).

$$K(x_i, x_j) = exp\left(-\gamma \left\| x_i - x_j \right\|^2\right) \qquad (4)$$

Operasi matematis pada kernel RBF ini menghitung eksponensial negatif dari perkalian antara parameter γ (Gamma) dengan kuadrat jarak *Euclidean* $\left\| \mathrm{x}_i - \mathrm{x}_j \right\|^2$. Kalkulasi jarak *Euclidean* tersebut berfungsi untuk mengukur tingkat kemiripan atau kedekatan spasial antara dua vektor fitur kalimat. Sementara itu, parameter γ bertindak sebagai pengontrol seberapa luas jangkauan pengaruh dari satu sampel data latih terhadap pembentukan garis batas keputusan, di mana nilai yang tinggi akan membuat batas keputusan secara ketat mengikuti pola data latih, sedangkan nilai yang rendah akan menghasilkan kurva pemisah yang lebih halus (*smooth*).

Untuk memastikan model SVM mencapai performa puncaknya dan terhindar dari kondisi *underfitting* atau *overfitting*, penelitian ini menerapkan tahapan *Hyperparameter Tuning* menggunakan metode *GridSearchCV* [28]. Model dievaluasi secara komputasional berulang kali melalui teknik validasi silang untuk mencari kombinasi nilai optimal pada ruang pencarian (*search space*) yang telah ditetapkan secara empiris. Ruang pencarian ini mencakup eksplorasi fungsi kernel (linear dan rbf),

rentang nilai parameter regularisasi atau C (0.1, 1, 10, dan 50), serta parameter γ (scale, 0.1, dan 1). Penetapan rentang metrik pengujian ini didasarkan pada pedoman optimasi heuristik standar untuk klasifikasi SVM, yang menyarankan eksplorasi nilai secara eksponensial (meliputi nilai rentang kecil, menengah, dan besar) untuk menemukan titik keseimbangan (*trade-off*) terbaik antara toleransi kesalahan (*soft margin*) dan margin pemisah yang tegas [29], [30]. Nilai C diuji dari batas regulasi kuat (C = 0.1) hingga penalti yang sangat ketat terhadap kesalahan latih (C = 50). Di sisi lain, opsi parameter γ difokuskan pada nilai desimal probabilitas rendah beserta fungsi kalkulasi otomatis scale. Pemilihan nilai batas ini sengaja diterapkan untuk mencegah kompleksitas komputasi yang berlebihan, mengingat matriks fitur teks dari ekstraksi TF-IDF memiliki sifat dimensi yang sangat renggang (*sparse matrix*) sehingga rentan mengalami *overfitting* jika diberikan nilai parameter pembobot yang terlalu agresif [31], [32].

2.6. Evaluasi Model

Evaluasi model merupakan tahapan kritis untuk mengukur seberapa baik algoritma *Support Vector Machine* (SVM) dalam menggeneralisasi data yang belum pernah dilihat sebelumnya. Evaluasi dalam penelitian ini dilakukan melalui dua tahapan utama: teknik validasi silang dan pengukuran matriks performa klasifikasi.

a. *Stratified 5-Fold Cross Validation*

Untuk menghindari bias evaluasi yang disebabkan oleh pembagian data latih dan data uji yang kebetulan menguntungkan atau merugikan model, penelitian ini menggunakan metode *Stratified 5-Fold Cross Validation*. Secara teknis, keseluruhan dataset dibagi secara acak menjadi 5 *fold* yang sama besar. Model kemudian dilatih sebanyak 5 iterasi [33]. Pada setiap iterasi, 4 *fold* digunakan sebagai data latih dan 1 *fold* sisanya digunakan sebagai data uji.

Penggunaan varian *Stratified* sangat esensial pada dataset ini karena ia menjamin bahwa rasio proporsi tiap kelas (Keluarga, Teman, Musuh) di setiap *fold* tetap seimbang dan merepresentasikan distribusi populasi data aslinya. Hal ini mencegah terjadinya imbalanced *fold*, di mana sebuah *fold* pengujian mungkin tidak memiliki sampel dari kelas minoritas sama sekali. Skor akhir model diambil dari nilai rata-rata performa selama 5 iterasi tersebut.

b. *Confusion Matrix*

Kinerja model dipetakan menggunakan *Confusion Matrix*, yaitu sebuah tabel matriks berukuran 3×3 (sesuai jumlah kelas) yang membandingkan label aktual (*ground truth*) dengan label hasil prediksi model [34]. Matriks ini memberikan representasi visual yang transparan mengenai di mana model sering melakukan kesalahan (*confusion*), misalnya seberapa sering interaksi "Keluarga" keliru diprediksi sebagai "Teman". Dari tabel *Confusion Matrix* tersebut, diekstrak empat komponen dasar matematika untuk evaluasi performa pada setiap kelasnya. Komponen pertama adalah *True Positive* (TP), yang merepresentasikan kondisi ketika model memprediksi kelas positif dan tebakan tersebut terbukti benar. Komponen kedua adalah *True Negative* (TN), yaitu kondisi ketika model dengan tepat memprediksi bahwa suatu data bukan bagian dari kelas target (negatif). Selanjutnya, terdapat komponen *False Positive* (FP) atau Kesalahan Tipe I, yang mendeskripsikan kegagalan model ketika memprediksi data sebagai kelas positif, padahal nilai aktualnya adalah negatif. Terakhir, matriks ini mengekstrak komponen *False Negative* (FN) atau Kesalahan Tipe II, yakni kondisi ketika model keliru memprediksi data sebagai kelas negatif, padahal secara aktual data tersebut bernilai positif.

c. Metrik Evaluasi Performa

Berdasarkan komponen TP, TN, FP, dan FN tersebut, kinerja model diukur menggunakan empat metrik utama evaluasi *Machine Learning yaitu* Akurasi (*Accuracy*), Presisi (*Precision*), Sensitivitas (*Recall*), dan F1-Score [35], [36], [37]:

- Akurasi (*Accuracy*): Merupakan rasio dari total prediksi yang benar (baik positif maupun negatif) terhadap keseluruhan sampel pengujian. Metrik ini menunjukkan tingkat kebenaran model secara umum.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \qquad (5)$$

p-ISSN: 2541-1179
e-ISSN: 2581-1711
JURNAL INSTEK
INFORMATIKA SAINS DAN TEKNOLOGI

- Presisi (*Precision*): Merupakan rasio ketepatan prediksi positif terhadap keseluruhan hasil yang diprediksi positif. Presisi tinggi menunjukkan bahwa ketika model memprediksi sebuah kalimat berelasi "Keluarga", model tersebut sangat dapat dipercaya.

$$Precision = \frac{TP}{TP+FP} \quad (6)$$

- Sensitivitas (*Recall*): Merupakan rasio keberhasilan model dalam menemukan semua sampel aktual yang benar-benar positif. *Recall* tinggi menunjukkan bahwa model berhasil mengenali hampir seluruh data "Keluarga" yang ada di dalam dataset tanpa ada yang terlewat.

$$Recall = \frac{TP}{TP+FN} \quad (7)$$

- *F1-Score*: Karena penelitian ini memiliki 3 kelas target (*multiclass*), mengandalkan Akurasi saja dapat menyesatkan jika terdapat ketidakseimbangan jumlah data antar kelas. Oleh karena itu, digunakan F1-Score yang merupakan rata-rata harmonik antara *Precision* dan *Recall*. Metrik ini memberikan bobot yang seimbang antara ketepatan prediksi dan sensitivitas pencarian.

$$F1\text{-}Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (8)$$

Dalam implementasi evaluasinya, nilai *Precision, Recall*, dan *F1-Score* dihitung untuk masing-masing kelas secara individual, kemudian dirata-ratakan menggunakan metode *Macro-Average* atau *Weighted-Average* untuk mendapatkan satu skor performa klasifikasi yang komprehensif.

3. HASIL DAN PEMBAHASAN

3.1. Hasil Pra-pemrosesan dan Ekstraksi Fitur

Tahap awal penelitian melibatkan pembersihan dan standarisasi data JSON. Dari total 637 segmen mentah yang telah dianotasi, proses deduplikasi mereduksi dataset menjadi 600 segmen unik. Setelah proses *masking* entitas tokoh menjadi ENTITY1 dan ENTITY2, sistem mengekstrak fitur teks berdasarkan arsitektur *Dual-Stream Pipeline*. Pada aliran pertama, implementasi TfidfVectorizer dengan kombinasi *N-gram* (1 hingga 3 kata) menghasilkan 11.987 fitur kosa kata pada teks dengan entitas tersamarkan. Aliran kedua menambahkan tiga skor leksikon yang merepresentasikan indikasi Keluarga, Teman, dan Musuh sebelum kedua aliran digabungkan melalui *FeatureUnion.*

3.2. Hasil Skenario *Hyperparameter Tuning*

Untuk mendapatkan batas pemisah (*hyperplane*) yang paling optimal, model *Support Vector Machine* (SVM) dieksplorasi menggunakan metode *GridSearchCV* dengan 5-*Fold Cross Validation*. Proses ini mengevaluasi 24 kombinasi kandidat (120 iterasi *fitting*) terhadap variasi fungsi *Kernel*, parameter regularisasi (C), dan parameter γ. Hasil optimasi komputasional menyimpulkan bahwa konfigurasi *hyperparameter* terbaik dicapai pada:

- Kernel: linear
- Nilai C: 10
- Nilai γ: scale

Terpilihnya kernel linear secara empiris relevan dengan karakteristik data teks. Aliran TF-IDF pada konfigurasi *N-gram* (1, 3) menghasilkan ruang fitur renggang (*sparse*) dengan 11.987 dimensi sebelum digabungkan dengan tiga fitur leksikon. Pada ruang berdimensi tinggi seperti ini, pemisahan linear dapat menjadi kompetitif tanpa transformasi non-linear yang kompleks seperti *Radial Basis Function* (RBF). Nilai C = 10 pada model usulan menunjukkan penalti yang relatif tegas terhadap kesalahan klasifikasi selama pelatihan untuk membentuk batas keputusan yang lebih presisi. Uji sensitivitas bobot leksikon menunjukkan bahwa B = 1,0; 1,5; 2,0; 2,5; dan 3,0 seluruhnya menghasilkan akurasi 0,750. Macro F1 berada pada rentang 0,7516–0,7518, sehingga selisih antarkandidat kurang dari 0,0002. B = 2,0 dipertahankan sebagai nilai tengah pada plateau akurasi tersebut karena peningkatan B hingga 2,5 atau 3,0 tidak memberikan peningkatan akurasi.

p-ISSN: 2541-1179
e-ISSN: 2581-1711
JURNAL INSTEK
INFORMATIKA SAINS DAN TEKNOLOGI

3.3. Perbandingan Model Usulan dengan Metode Pembanding

Perbandingan dilakukan untuk menilai kontribusi aliran leksikon dan mengukur dampak *masking* secara empiris. Hasil pada Tabel 2 menunjukkan bahwa model *Dual-Stream* memperoleh akurasi 0,750 dan *Macro F1* 0,752. Ketika aliran leksikon dihilangkan dan model hanya menerima TF-IDF dari teks dengan entitas tersamarkan, akurasi turun menjadi 0,715 dan *Macro F1* menjadi 0,715. Dengan demikian, keberadaan aliran leksikon meningkatkan akurasi sebesar 3,5 poin persentase dibandingkan *baseline* teks dengan entitas tersamarkan dan memulihkan sebagian informasi relasional yang berkurang setelah *masking.*

Tabel 2. Hasil Perbandingan Kinerja Model Usulan dan Metode Pembanding

Metode	Representasi	*Accuracy*	*Macro Precision*	*Macro Recall*	*Macro F1*
Dual-Stream (Usulan)	TF-IDF entitas tersamarkan + leksikon teks asli	0.750	0.757	0.750	0.752
TF-IDF Entitas Tersamarkan + SVM	TF-IDF teks dengan entitas tersamarkan	0.715	0.715	0.715	0.715
TF-IDF Teks Asli + SVM	TF-IDF teks asli tanpa masking	0.837	0.837	0.837	0.837

Baseline TF-IDF pada teks asli menghasilkan akurasi tertinggi, yaitu 0,837, atau 8,7 poin persentase di atas model usulan. Hasil ini menunjukkan bahwa nama tokoh dan leksikon asli mengandung sinyal prediktif yang kuat pada skenario *Stratified 5-Fold Cross Validation* tingkat kalimat. Namun, karena pembagian *fold* dilakukan pada tingkat sampel, tokoh yang sama dapat muncul pada data latih dan data uji sehingga *baseline* tanpa masking berpotensi memanfaatkan pola spesifik identitas selain bukti relasional. Oleh sebab itu, hasil pembandingan tidak digunakan untuk menyatakan bahwa *Dual-Stream* selalu lebih unggul daripada teks asli. Temuan ini justru menempatkan model usulan sebagai kompromi antara pengurangan ketergantungan terhadap identitas tokoh dan pemulihan sebagian informasi leksikal yang hilang akibat masking.

Jika dibandingkan langsung, *masking* tanpa aliran leksikon menurunkan akurasi sebesar 12,2 poin persentase dari *baseline* teks asli (83,7% menjadi 71,5%), sedangkan *Dual-Stream* memulihkan 3,5 poin persentase menjadi 75,0%. Peningkatan setelah penambahan fitur leksikon konsisten dengan penelitian berbasis SVM dan sumber leksikal yang melaporkan manfaat fitur leksikon sebagai informasi tambahan [8], [9]. Sementara itu, penurunan informasi setelah *entity masking* sejalan dengan Alt dkk. [14], yang menunjukkan bahwa *masking* dapat membantu tugas ekstraksi relasi tetapi mengurangi informasi terkait entitas. Besaran peningkatan tidak dibandingkan secara langsung karena dataset dan tugas pada penelitian terdahulu berbeda.

3.4. Evaluasi Kinerja Klasifikasi

Kinerja klasifikasi model Support Vector Machine (SVM) dengan arsitektur Dual-Stream dievaluasi menggunakan 600 prediksi hasil *Stratified 5-Fold Cross Validation*. Rangkuman metrik evaluasi yang terdiri dari *Precision, Recall,* dan *F1-Score* untuk masing-masing kelas disajikan pada Tabel 3.

p-ISSN: 2541-1179
e-ISSN: 2581-1711
JURNAL INSTEK
INFORMATIKA SAINS DAN TEKNOLOGI

Tabel 3. Hasil Pengujian Performa Model Klasifikasi

Kelas Relasi	*Precision*	*Recall*	*F1-Score*	Jumlah Sampel (*Support*)
Keluarga	0.87	0.75	0.80	200
Musuh	0.71	0.78	0.74	200
Teman	0.70	0.72	0.71	200
Rata-rata (*Macro Avg*)	0.76	0.75	0.75	600
Akurasi Keseluruhan	-	-	0.75 (75%)	600

Secara keseluruhan, model usulan mencapai akurasi 75% dengan Macro Precision sebesar 0,76, Macro Recall 0,75, dan Macro F1 sekitar 0,75. Nilai tersebut berada jauh di atas peluang tebakan acak pada klasifikasi tiga kelas yang seimbang. Namun, interpretasi performa tidak hanya didasarkan pada akurasi absolut, melainkan juga pada hasil perbandingan *baseline* pada Tabel 2 yang menunjukkan adanya *trade-off* antara masking, retensi informasi leksikal, dan potensi ketergantungan terhadap identitas tokoh.

Analisis per kelas menunjukkan variasi performa yang berkaitan dengan karakteristik leksikal setiap relasi. Kelas Keluarga memperoleh *Precision* tertinggi sebesar 87% dan *F1-Score* 80%. Kinerja ini konsisten dengan keberadaan kosakata kekerabatan bahasa Bali seperti meme, bapa, pianak, cucu, dan kurenan yang memberikan sinyal diskriminatif kuat pada aliran leksikon. Dibandingkan *baseline* teks dengan entitas tersamarkan, peningkatan performa model usulan menunjukkan bahwa pemanfaatan leksikon asli memberikan kontribusi terukur, meskipun belum cukup untuk melampaui *baseline* teks asli tanpa *masking*.

Di sisi lain, kelas Musuh (*F1-Score* 0.74) dan Teman (*F1-Score* 0.71) menunjukkan kinerja yang moderat namun stabil. Kesulitan algoritma dalam mendorong akurasi kedua kelas ini melampaui 80% mengindikasikan adanya batasan representasi *N-gram* dalam memahami konteks kalimat naratif yang kompleks, yang akan dibedah lebih lanjut melalui analisis *confusion matrix*.

3.5. Analisis Kesalahan Model Berdasarkan *Confusion Matrix*

Untuk membedah rincian pola misklasifikasi (*misclassification*) pada masing-masing kelas secara lebih granular, *confusion matrix* dari hasil pengujian divisualisasikan pada Tabel 4.

Tabel 4. Confusion Matrix Model Usulan

	Prediksi: Keluarga	**Prediksi: Musuh**	**Prediksi: Teman**
Asli: Keluarga	150	19	31
Asli: Musuh	13	155	32
Asli: Teman	10	45	145

Berdasarkan distribusi nilai pada Tabel 4, diperoleh beberapa temuan signifikan yang merepresentasikan interaksi antara algoritma komputasi dan linguistik *Satua Bali*:

a. Dominasi *True Positive* (Ketahanan Model)

Konsentrasi nilai tertinggi berada pada diagonal utama matriks, yang merepresentasikan prediksi benar sebanyak 150 untuk Keluarga, 155 untuk Musuh, dan 145 untuk Teman. Pola ini menunjukkan bahwa batas keputusan SVM mampu memisahkan sebagian besar sampel pada ruang fitur gabungan berdimensi tinggi, meskipun masih terdapat tumpang tindih yang cukup besar antara kelas Teman dan Musuh.

b. Ambiguitas Leksikal antara "Teman" dan "Musuh" (Titik Lemah Model)

Matriks menunjukkan kebingungan terbesar antara kelas Teman dan Musuh: 45 sampel Teman diprediksi sebagai Musuh dan 32 sampel Musuh diprediksi sebagai Teman. Contoh Teman→Musuh adalah "Gelisang satua enggal, majalan kone ajaka dadua ngliwatin tukad cengcenge, I Lutung magandong ditundun I Kekuane." Segmen ini menggambarkan tindakan kooperatif, tetapi tidak memuat penanda leksikal pertemanan yang eksplisit sehingga pola *N-gram* lokal tidak cukup kuat. Contoh

Musuh→Teman adalah "I Lutung ngaturang pituduh teken Jero Dukuh mangda nuntun ikuhne apang olahane nenten pahit." Secara konteks cerita, tuturan tersebut merupakan bagian dari strategi penipuan, tetapi secara lokal mengandung bentuk ujaran yang menyerupai pemberian saran. Kedua contoh tersebut menunjukkan bahwa representasi tingkat kalimat belum menangkap urutan peristiwa dan maksud pragmatis yang diperlukan untuk membedakan interaksi tulus dari manipulasi.

c. Isolasi Semantik pada Kelas "Keluarga"

Bertolak belakang dengan ambiguitas tersebut, kelas Keluarga menunjukkan tingkat pemisahan leksikal yang relatif lebih kuat. Kesalahan antara Keluarga dan Teman lebih rendah dibandingkan pertukaran Teman dan Musuh. Temuan ini menunjukkan bahwa penanda kekerabatan dalam bahasa Bali cenderung lebih eksplisit dan lebih mudah direpresentasikan sebagai fitur diskriminatif dibandingkan relasi pertemanan atau permusuhan yang sering bergantung pada konteks tindakan dan urutan peristiwa.

3.6. Keterbatasan Representasi Relasi Tingkat Kalimat

Penelitian ini merepresentasikan relasi pada tingkat kalimat atau segmen lokal agar setiap sampel memiliki satu pasangan tokoh dan bukti interaksi yang jelas. Pembatasan ini menjaga konsistensi anotasi serta memungkinkan pembandingan TF-IDF dan leksikon secara terkontrol tanpa menerapkan resolusi koreferensi lintas kalimat yang berada di luar cakupan eksperimen.

Meskipun demikian, relasi dalam teks naratif tidak selalu dapat disimpulkan dari satu kalimat. Status hubungan dapat dipengaruhi oleh konteks antarkalimat, rujukan pronomina atau koreferensi, urutan peristiwa, serta perubahan relasi sepanjang cerita. Tokoh yang secara global bermusuhan, misalnya, dapat menggunakan tuturan yang tampak ramah pada satu kalimat sebagai bagian dari strategi penipuan. Oleh karena itu, prediksi model pada penelitian ini perlu dipahami sebagai klasifikasi bukti relasi lokal, bukan estimasi keadaan hubungan global antar tokoh. Keterbatasan ini juga berimplikasi pada evaluasi: pembagian acak tingkat kalimat memungkinkan nama tokoh yang sama muncul pada *fold* latih dan uji, sehingga performa *baseline* teks asli berpotensi lebih tinggi akibat sinyal identitas. Penelitian lanjutan perlu mengevaluasi jendela multikalimat, resolusi koreferensi, pemodelan urutan peristiwa, serta skema *entity-disjoint* atau *story-disjoint split* untuk menguji generalisasi terhadap tokoh dan cerita yang benar-benar baru.

4. KESIMPULAN

Penelitian ini merancang model ekstraksi relasi karakter *Satua Bali* untuk tiga kelas, yaitu Keluarga, Teman, dan Musuh, menggunakan SVM dengan arsitektur *Dual-Stream Pipeline*. Kontribusi data terletak pada proses kurasi dan skema anotasi relasi, bukan pada teks mentah yang berasal dari sumber web publik. Pada 600 segmen unik, model usulan mencapai akurasi 75,0% dan *Macro F1* 75,2%. TF-IDF pada teks dengan entitas tersamarkan tanpa aliran leksikon memperoleh akurasi 71,5%, sehingga aliran leksikon memulihkan 3,5 poin persentase, sedangkan TF-IDF pada teks asli mencapai 83,7% dan menunjukkan kuatnya sinyal identitas pada evaluasi tingkat kalimat. Dengan demikian, *Dual-Stream* diposisikan sebagai strategi mitigasi *trade-off masking*, bukan penyelesaian definitif. Keterbatasan utama penelitian adalah belum dimodelkannya koreferensi, konteks antarkalimat, dan urutan peristiwa. Penelitian lanjutan direkomendasikan menggunakan konteks multikalimat dan *Entity-Disjoint Split* agar tokoh pada data uji tidak muncul pada data latih; *Story-Disjoint Split* dapat digunakan sebagai evaluasi tambahan untuk menguji generalisasi terhadap cerita baru.

5. DAFTAR PUSTAKA

[1] Komang Ria Ariestiyani, I Made Suweta, I Putu Suardipa, and I Ketut Ngurah Ardiawan, "Development Of A Flipbook Of Satua Bali I Lutung Teken I Kakua With Mathematical Numeracy Literacy In Grade Iii Elementary School," *IJPSE Indones. J. Prim. Sci. Educ.*, vol. 6, no. 1 SE-Articles, pp. 62–70, Nov. 2025, doi: 10.33752/ijpse.v6i1.9222.

[2] N. Made, A. Arsita, N. T. Astawa, U. H. Negeri, G. Bagus, and S. Denpasar, "Suaraku Bahasaku: Dinamika Kesulitan Berbahasa Bali dalam Dunia Siswa SMA Negeri 2 Mengwi," *J. Onoma Pendidikan, Bahasa, dan Sastra*, vol. 12, no. 2 SE-Articles, pp. 1525–1547, 2026, doi: 10.30605/khw7f008.

[3] X. Liu, C. Hu, R. Zhang, J. Chen, and B. Xu, "Improving Data Annotation for Low-Resource Relation Extraction with Logical Rule-Augmented Collaborative Language Models," in *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, L. Chiruzzo, A. Ritter, and L. Wang, Eds., Albuquerque, New Mexico: Association for Computational Linguistics, Apr. 2025, pp. 1497–1510. doi: 10.18653/v1/2025.naacl-long.70.

[4] P. Agustina and R. A. Putri, "Combination of Lexical Resources and Support Vector Machine for Film Sentiment Analysis," *Sink. J. dan Penelit. Tek. Inform.*, vol. 8, no. 3 SE-, pp. 1526–1538, Jul. 2024, doi: 10.33395/sinkron.v8i3.13733.

[5] A. R. Noor and D. Normawati, "Perbandingan support vector machine dan k-nearest neighbor pada analisis sentimen pelaksanaan pemilihan presiden tahun 2024," *J. Ilm. Penelit. dan Pembelajaran Inform.*, vol. 11, no. 2, pp. 1021–1030, 2026.

[6] R. Difandana, I. Imaduddin, and I. Indra, "Comparative Analysis of Naive Bayes and Support Vector Machine for Hate Speech Classification," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 6, no. 1, pp. 414–422, 2026, doi: 10.57152/malcom.v6i1.2571.

[7] M. N. Pratama and D. P. Ardianto, "Klasifikasi Pesan SMS Menggunakan Metode TF-IDF dan Support Vector Machine," *Semin. Nas. Teknol. Sains*, vol. 5, no. 1 SE-Articles, pp. 151–160, 2026, doi: 10.29407/9y2vn411.

[8] V. Nurcahyawati and Z. Mustaffa, "Vader Lexicon and Support Vector Machine Algorithm to Detect Customer Sentiment Orientation," *J. Inf. Syst. Eng. Bus. Intell.*, vol. 9, no. 1 SE-Articles, pp. 108–118, Apr. 2023, doi: 10.20473/jisebi.9.1.108-118.

[9] R. H. Muhammadi, T. G. Laksana, and A. B. Arifa, "Combination of Support Vector Machine and Lexicon- Based Algorithm in Twitter Sentiment Analysis," *Khazanah Inform. J. Ilmu Komput. dan Inform.*, pp. 59–71, 2022.

[10] A. A. Nabil, F. R. Wildantama, D. Satrianto, M. G. Bakara, I. Budiawan, and D. Mulyati, "Klasifikasi Hoax Menggunakan Metode TF-IDF + SVM: Penelitian," *J. Pengabdi. Masy. dan Ris. Pendidik.*, vol. 4, no. 3 SE-Articles, pp. 17653–17660, Jan. 2025, doi: 10.31004/jerkin.v4i3.5078.

[11] F. Koto, A. Rahimi, J. H. Lau, and T. Baldwin, "IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP," in *Proceedings of the 28th International Conference on Computational Linguistics*, D. Scott, N. Bel, and C. Zong, Eds., Barcelona, Spain (Online): International Committee on Computational Linguistics, Dec. 2020, pp. 757–770. doi: 10.18653/v1/2020.coling-main.66.

[12] W. Wongso, D. S. Setiawan, S. Limcorn, and A. Joyoadikusumo, "NusaBERT: Teaching IndoBERT to be Multilingual and Multicultural," in *Proceedings of the Second Workshop in South {E}ast {A}sian Language Processing*, D. Wijaya, A. F. Aji, C. Vania, G. I. Winata, and A. Purwarianti, Eds., Online: Association for Computational Linguistics, Jan. 2025, pp. 10–26. [Online]. Available: https://aclanthology.org/2025.sealp-1.2/

[13] N. Fajriyah, N. T. Lapatta, D. W. Nugraha, and R. Laila, "Implementasi svm dan smote pada analisis sentimen media sosial x terhadap pelantikan agus harimurti yudhoyono," *J. Ilm. Penelit. dan Pembelajaran Inform.*, vol. 10, no. 2, pp. 1359–1370, 2025.

[14] C. Alt, A. Gabryszak, and L. Hennig, "Probing Linguistic Features of Sentence-Level Representations in Neural Relation Extraction," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, D. Jurafsky, J. Chai, N. Schluter, and J. Tetreault, Eds., Online: Association for Computational Linguistics, Jul. 2020, pp. 1534–1545. doi: 10.18653/v1/2020.acl-main.140.

[15] E. Pirade, I. K. Gede, D. Putra, D. Purnami, and S. Putri, "Automated Generation of Folklore Short Stories Using T5 Transformer Model," *J. Appl. Informatics Comput.*, vol. 9, no. 5, pp. 2338–2346, 2025.

[16] K. Lee, D. Eck, D. Ippolito, A. Nystrom, C. Callison-burch, and N. Carlini, "Deduplicating Training Data Makes Language Models Better," *Proc. 60th Annu. Meet. Assoc. Comput. Linguist.*, 2021.

[17] S. Martínez, D. Sánchez, A. Valls, and M. Batet, "Privacy protection of textual attributes through a semantic-based masking method," *Inf. Fusion*, vol. 13, no. 4, pp. 304–314, 2012, doi: https://doi.org/10.1016/j.inffus.2011.03.004.

[18] G. Park, J. Rayz, and L. Pouchard, "Figure Descriptive Text Extraction using Ontological Representation," 2009.

[19] G. M. Emelyanov, D. V Mikhailov, and A. P. Kozlov, "The TF-IDF measure and analysis of links between words within N-grams in the formation of knowledge units for open tests," *Pattern Recognit. Image Anal.*, vol. 27, no. 4, pp. 825–831, 2017, doi: 10.1134/S1054661817040058.

[20] J. Hou, X. Li, R. Zhu, C. Zhu, Z. Wei, and C. Zhang, "A Neural Relation Extraction Model for Distant Supervision in Counter-Terrorism Scenario," *IEEE Access*, vol. 8, pp. 225088–225096, 2020, doi: 10.1109/ACCESS.2020.3042672.

[21] Z. Yang, Z. Li, Z. Xu, Z. Gan, and W. Cao, "A Joint Relation Extraction Model Based on Domain N-Gram Adapter and Axial Attention for Military Domain," in *Web Information Systems and Applications*, L. Yuan, S. Yang, R. Li, E. Kanoulas, and X. Zhao, Eds., Singapore: Springer Nature Singapore, 2023, pp. 233–245.

[22] I. M. Wahyu, G. Negara, N. Agus, and S. Er, "Basic Word Extraction Algorithm Based on Morphological Rules for Balinese Texts," vol. 8, no. 4, pp. 401–410, 2020.

[23] A. Nurkasanah and M. Hayaty, "Feature Extraction using Lexicon on the Emotion Recognition Dataset of Indonesian Text," *Ultim. J. Tek. Inform.*, vol. 14, no. 1, pp. 20–27, 2022, doi: 10.31937/ti.v14i1.2540.

[24] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede, "Lexicon-basedmethods for sentiment analysis," *Comput. Linguist.*, vol. 37, no. 2, pp. 267–307, 2011, doi: 10.1162/COLI_a_00049.

[25] A. Siddiqua, V. Bindumathi, G. Raghu, and Y. S. Vamsi Bhargav, "Aspect-based Sentiment Analysis (ABSA) using Machine Learning Algorithms," in *2024 Third International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE)*, 2024, pp. 1–6. doi: 10.1109/ICDCECE60827.2024.10549140.

[26] E. A. Kyei *et al.*, "A Hybrid Sentiment Analysis Model to Detect Racist Tweets Using Lexicon-Based Sentiment Analysis and a Support Vector Machine Algorithm," in *AI Revolution: Research, Ethics and Society*, H. R. Arabnia, L. Deligiannidis, S. Amirian, F. Ghareh Mohammadi, and F. Shenavarmasouleh, Eds., Cham: Springer Nature Switzerland, 2026, pp. 365–389.

[27] M. Mittal, H. M. Al-Jawahry, N. Varshney, S. P. Kumar, J. Michaelson, and R. A. Reddy, "Improving Support Vector Machine Performance with Advanced Kernel Methods," in *2024 7th International Conference on Contemporary Computing and Informatics (IC3I)*, 2024, pp. 1749–1754. doi: 10.1109/IC3I61595.2024.10828664.

[28] M. Li, Y. Liu, and J. Wang, "A new parameter optimization algorithm of SVM," in *2011 International Conference on Advanced Intelligence and Awareness Internet (AIAI 2011)*, 2011, pp. 174–178. doi: 10.1049/cp.2011.1451.

[29] J. Rusman, B. Z. Haryati, and A. Michael, "Optimisasi Hiperparameter Tuning pada Metode Support Vector Machine untuk Klasifikasi Tingkat Kematangan Buah Kopi," *J. Komput. dan Inform.*, vol. 11, no. 2, pp. 195–202, 2023, doi: 10.35508/jicon.v11i2.12571.

[30] Siti Kania Nur Alya Putri, Indah Jumiatin, Indri Sulistia, Novan Alkaf Bahraini Saputra, and Nuruddin Wiranda, "Implementation of Hyperparameter Tuning for Classification Models in Heart Disease Risk Prediction," vol. 5, no. October, pp. 1181–1189, 2025.

[31] B. Astario, T. Tukino, A. Hananto, F. Nurapriani, and E. Novalia, "Text Data Classification Using the SVM Model on the LMDB Minecraft Dataset," *J. Inf. dan Teknol.*, pp. 21–26, May 2025, doi: 10.60083/jidt.vi0.620.

[32] I. G. M. N. A. Bhawanaputra, I. G. I. Sudipa, N. P. S. Meinarni, I. G. A. A. M. Aristamy, and I. Pratistha, "SVM-Based Pediatric Disease Classification Model from the Balinese Lontar Usada Rare Manuscript," *Sink. J. dan Penelit. Tek. Inform.*, vol. 10, no. 1, pp. 481–494, 2026, doi:

10.33395/sinkron.v10i1.15508.

[33] V. W. Lumumba, D. Kiprotich, M. L. Mpaine, N. G. Makena, and M. D. Kavita, "Comparative Analysis of Cross-Validation Techniques: LOOCV, K-folds Cross-Validation, and Repeated K-folds Cross-Validation in Machine Learning Models," *Am. J. Theor. Appl. Stat.*, vol. 13, no. 5, pp. 127–137, 2024, doi: 10.11648/j.ajtas.20241305.13.

[34] I. Markoulidakis and G. Markoulidakis, "Probabilistic Confusion Matrix: A Novel Method for Machine Learning Algorithm Generalized Performance Analysis," *Technologies*, vol. 12, no. 7, 2024, doi: 10.3390/technologies12070113.

[35] M. Gong, "A Novel Performance Measure for Machine," vol. 13, no. 1, pp. 1–19, 2021, doi: 10.5121/ijmit.2021.13101.

[36] G. Naidu, T. Zuva, and E. M. Sibanda, "A Review of Evaluation Metrics in Machine Learning Algorithms," in *Artificial Intelligence Application in Networks and Systems*, R. Silhavy and P. Silhavy, Eds., Cham: Springer International Publishing, 2023, pp. 15–25.

[37] A. Tilevik, "Classification and Performance Metrics," in *Multivariate Statistics and Machine Learning in R For Beginners: With Applications in Biology and Medicine*, Cham: Springer Nature Switzerland, 2025, pp. 147–170. doi: 10.1007/978-3-032-01851-9_10.