BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 4, June 2026 | Hal 1666-1674
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i4.1219

Analisis Efektivitas IndoBERT untuk Klasifikasi Multilabel Terjemahan Hadis Bukhari Menggunakan Logistic Regression

Achmad Yamin Harahap*, Nazruddin Safaat H, Surya Agustian, Suwanto Sanjaya, Teddie D

Fakultas Sains dan Teknologi, Teknik Informatika, Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Indonesia
Email: [1]111950111665@students.uin-suska.ac.id, [2]nazruddin.safaat@uin-suska.ac.id, [3,*]surya.agustian@uin-suska.ac.id,
[4]suwantosanjaya@uin-suska.ac.id, [5]teddie.darmizal@uin-suska.ac.id
Email Penulis Korespondensi: surya.agustian@uin-suska.ac.id

Abstrak−Hadis yaitu sumber ajaran kedua setelah Al-Qur'an yang memandu umat Islam dalam menghadapi berbagai aspek kehidupan. Salah satu hadis yang terkenal yaitu hadis Shahih Al-Bukhari. Kandungan makna yang kompleks dan sering kali memuat lebih dari satu kategori pesan menjadi tantangan utama dalam proses pengelompokan teks secara manual, terutama ketika jumlah data semakin besar. Pada penelitian ini, makna hadis seringkali memiliki lebih dari satu kandungan pesan, seperti anjuran, larangan, dan informasi. Penelitian ini bertujuan untuk mengevaluasi sistem klasifikasi otomatis pada hadis Shahih Bukhari terjemahan Bahasa Indonesia ke dalam tiga kategori, yaitu Informasi, Anjuran, dan Larangan. Penelitian ini dilatarbelakangi oleh banyaknya jumlah hadis yang menyebabkan masyarakat memerlukan waktu dan pemahaman yang lebih mendalam untuk mengetahui kandungan utama dari setiap hadis. Dengan adanya sistem klasifikasi ini, diharapkan pengguna dapat lebih mudah mengidentifikasi pesan utama yang terkandung dalam hadis sehingga proses pencarian, pembelajaran, dan pemahaman hadis dapat dilakukan secara lebih efektif dan efisien.. IndoBERT digunakan untuk menghasilkan representasi vektor kontekstual yang mampu menangkap makna semantik teks secara lebih mendalam, sementara Logistic Regression dipilih karena efisiensi dan stabilitasnya pada data berdimensi tinggi. Evaluasi dilakukan menggunakan pendekatan pembagian data train, validasi, dan pengujian, serta metrik akurasi dan macro F1-score. Penelitian ini meberikan rata-rata F1-score hasil yaitu sebesar 67,43%, menunjukkan bahwa kombinasi IndoBERT dan Logistic Regression dapat menghasilkan performa klasifikasi yang baik dan konsisten pada tugas multilabel.

Kata Kunci: Hadis Bukhari; Klasifikasi; IndoBERT; Logistic Regression

Abstract−Hadith serves as the second source of guidance after the Quran, directing Muslims in various aspects of life; the *Sahih al-Bukhari* collection is among the most renowned. The complex nature of their meanings often encompassing multiple categories of messages poses a significant challenge for manual text classification, particularly as data volume grows. In this study, the content of the hadith often includes multiple message types, such as recommendations, prohibitions, and general information. This research aims to evaluate an automated classification system for Indonesian translations of *Sahih al-Bukhari* hadith, categorizing them into three classes: Information, Recommendation, and Prohibition. The study is motivated by the vast number of hadith, which requires significant time and deep understanding for people to grasp the core message of each one. This classification system is intended to facilitate the identification of primary messages, thereby making the processes of searching, studying, and understanding hadith more effective and efficient. IndoBERT is employed to generate contextual vector representations capable of capturing deeper semantic meaning, while Logistic Regression is selected for its efficiency and stability with high-dimensional data. Evaluation is conducted using a train-validation-test split approach, alongside accuracy and macro F1-score metrics. The study achieved an average F1-score of 67.43%, demonstrating that the combination of IndoBERT and Logistic Regression yields strong, consistent classification performance for this multi-label task.

Keywords: Bukhari Hadith; Clasification; IndoBERT; Multilabel; Logistic Regression

1. PENDAHULUAN

Hadis merupakan salah satu sumber ajaran dan hukum Islam yang sangat penting setelah Al-Qur'an, berisi perkataan, perbuatan, dan persetujuan Nabi Muhammad SAW yang menjadi pedoman hidup bagi umat Islam di berbagai aspek kehidupan [1]. Karena karakteristiknya yang kaya makna dan padat informasi, hadis sering kali berisi lebih dari satu jenis pesan, seperti anjuran, larangan, dan informasi deskriptif, yang membuat pemrosesan secara manual menjadi kompleks dan memakan waktu. Digitalisasi dan volume teks hadis yang terus bertambah mendorong berkembangnya penggunaan Natural Language Processing (NLP) untuk otomatisasi analisis teks keagamaan ini dalam skala besar [2].

Dalam domain NLP, tugas text clasification khususnya multilabel classification akan menjadi tantangan penting karena setiap contoh teks bisa memiliki lebih dari satu label yang relevan secara semantik yang lebih baik dibandingkan model generik lainnya, sehingga cocok digunakan sebagai feature extraktor dalam berbagai tugas klasifikasi teks berbahasa Indonesia. Model ini didasarkan pada arsitektur Transformer yang dapat menangkap hubungan konteks dua arah antar token secara efektif, sebagaimana diperkenalkan dalam penelitian sebelumnya tentang model pre-trained Bahasa Indonesia yang telah menunjukkan peningkatan performa pada beberapa tugas NLP seperti analisis sentimen dan klasifikasi teks [3].

Berbagai penelitian terdahulu telah mengeksplorasi pendekatan klasifikasi teks Hadis Bukhari terjemahan menggunakan metode machine learning dan deep learning yang beragam. Misalnya, pendekatan multilabel Support Vektor Machine (SVM) dan Long Short Term Memory (LSTM), menunjukkan bahwa LSTM mengungguli SVM pada label Larangan dan Anjuran dengan nilai F1-score masing-masing sebesar 82,66% dan 71,04% [4]. Sebaliknya, SVM menunjukkan performa lebih baik pada label Informasi dengan F1-score sebesar 67,40%. Selain itu penelitian lain juga membandingkan antara metode Random Forest (RF) dengan Long Short Term Memory (LSTM), yang mana LSTM menunjukkan keunggulan pada label Anjuran dan Informasi dengan nilai F1-score masing-masing sebesar 69,99% dan

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 4, June 2026 | Hal 1666-1674
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i4.1219

64,27% [5]. Sebaliknya, RF menunjukkan performa lebih baik pada label larangan dengan F1-score 84,45%. Kemudian penelitian lainnya menggunakan metode Logistic Regression (LR) dan ekstraksi fitur menggunakan metode TF-IDF (Term Frequency–Inverse Document Frequency), yang mampu menghasilkan F1-Score sebesar 73,95% pada kategori Anjuran, 79,86% pada kategori Larangan, dan 57,23% pada kategori Informasi [6].

Meskipun beberapa penelitian sebelumnya menunjukkan bahwa metode berbasis TF-IDF mampu menghasilkan performa yang kompetitif, khususnya ketika dikombinasikan dengan algoritma seperti SVM, Random Forest, atau Logistic Regression, pendekatan tersebut umumnya mengandalkan representasi statistik kata yang kurang memperhatikan konteks kalimat. Sebaliknya, IndoBERT menghasilkan representasi kontekstual yang mampu menangkap hubungan semantik antar kata dalam teks hadis. Oleh karena itu, penelitian ini bertujuan untuk mengevaluasi sejauh mana representasi fitur IndoBERT dapat dimanfaatkan oleh Logistic Regression pada tugas klasifikasi multilabel terjemahan hadis Bukhari.

Penggunaan IndoBERT sebagai *feature extractor* pada penelitian ini didasarkan pada kemampuannya dalam menghasilkan representasi teks yang mempertimbangkan konteks penggunaan kata dalam kalimat bahasa Indonesia. Meskipun beberapa penelitian sebelumnya menunjukkan bahwa pendekatan berbasis TF-IDF masih mampu memberikan hasil yang lebih tinggi pada kategori tertentu, metode tersebut umumnya merepresentasikan teks berdasarkan frekuensi kemunculan kata sehingga informasi konteks antar kata belum sepenuhnya terakomodasi. Pada kasus klasifikasi teks hadis, kondisi ini berpotensi menimbulkan kesulitan dalam membedakan hadis yang memiliki kosakata serupa tetapi mengandung maksud yang berbeda, terutama pada kategori tertentu.

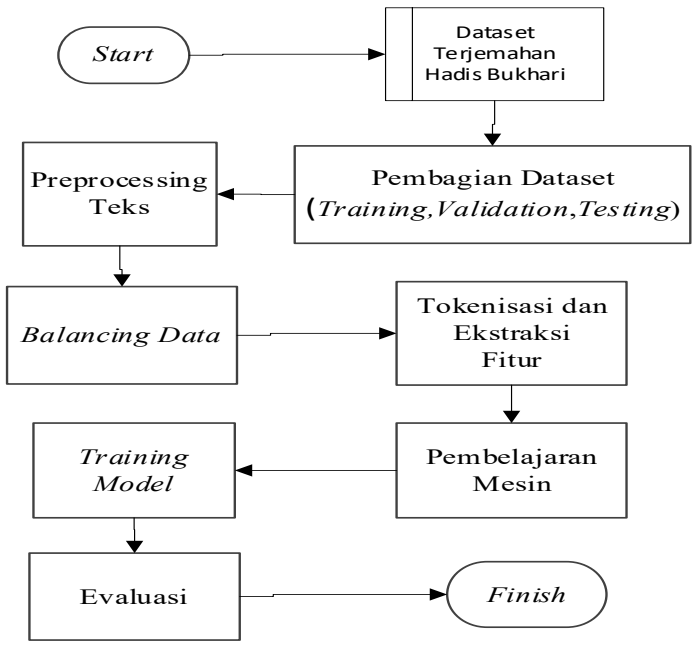
Penelitian ini tidak bertujuan untuk membuktikan bahwa IndoBERT lebih unggul dibandingkan metode tradisional atau pendekatan *fine-tuning* Transformer secara *end-to-end*. Sebaliknya, penelitian ini berfokus pada evaluasi efektivitas representasi fitur yang dihasilkan IndoBERT ketika dikombinasikan dengan Logistic Regression sebagai algoritma klasifikasi. Logistic Regression dipilih karena memiliki kompleksitas yang relatif rendah, mudah diimplementasikan, serta banyak digunakan sebagai *baseline classifier* dalam penelitian klasifikasi teks. Dengan pendekatan ini, kualitas representasi fitur IndoBERT dapat dievaluasi secara lebih objektif tanpa melibatkan proses pelatihan ulang model Transformer secara penuh.

Melalui kombinasi tersebut, penelitian ini berupaya mengkaji sejauh mana representasi fitur IndoBERT dapat membantu proses klasifikasi multilabel pada terjemahan hadis Bukhari, khususnya pada kategori Anjuran, Larangan, dan Informasi. Hasil penelitian kemudian dibandingkan dengan metode yang telah digunakan pada penelitian sebelumnya untuk melihat kelebihan dan keterbatasan pendekatan yang diusulkan.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Penelitian ini dilakukan dengan dasar Flowchart yang digambarkan pada Gambar 1.



Gambar 1. Alur Metode Penelitian

Pada Gambar 1 memvisualisasikan alur dari penelitian yang dilakukan. Berikut ini penjelasan lebih rinci mengenai setiap langkah dari flowchart berikut.

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 4, June 2026 | Hal 1666-1674
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i4.1219

2.2 *Dataset*

Dalam penelitian ini, dataset yang digunakan yaitu kumpulan hadis Bukhari terjemahan bahasa Indonesia. Data ini diperoleh dari penelitian terdahulu [4], [5]yang telah melalui proses pelabelan oleh para ahli. Dataset yang digunakan terdiri dari 7.000 entri hadis yang telah dilabeli dalam tiga kategori utama, yaitu Anjuran, Informasi, dan Larangan. Ketiga kategori ini diambil dari penelitian terdahulu yang menjadi dasar dari penelitian ini. Berikut representasi struktur dataset dengan format multilabel yang ditunjukkan pada tabel 1.

Tabel 1. Representasi Dataset

No.	Hadis	Kelas		
		Anjuran	Larangan	Informasi
6781	Telah menceritakan kepada kami [Musaddad] telah menceritakan kepada kami [Abdul Warits] dari [Abdul Aziz] dari [Anas] dia berkata; Nabi shallallahu 'alaihi wasallam melarang seorang laki-laki memakai minyak za'faran.	0	1	1
6931	Telah menceritakan kepada kami [Qutaibah bin Sa'id] dari [Malik] dari [Nafi'] dari [Ibnu Umar], Rasulullah Shallallahu'alaihiwasallam melarang (jual beli) najasy (penipuan).	0	1	0
6945	Telah bercerita kepada kami ['Abdullah bin Yusuf] telah mengabarkan kepada kami [Malik] dari [Nafi'] dari ['Abdullah bin 'Umar radliallahu 'anhuma] bahwa Rasulullah shallallahu 'alaihi wasallam memerintahkan untuk membunuh anjing.	1	0	1

Dalam Tabel 1 merupakan pelabelan klasifikasi multilabel, yang artinya setiap hadis tidak dibatasi hanya pada satu kategori. Nilai 1 berarti label tersebut ada pada hadis, dan nilai 0 berarti tidak ada label pada hadis.Maka dari itu digunakannya skema transformasi label One-vs-Rest. Karena secara kontekstual, skema OvR selaras dengan karakteristik teks hadis, di mana setiap kategori hukum (seperti larangan atau anjuran) diidentifikasi berdasarkan keberadaan fitur semantik spesifiknya secara mandiri (*independent binary concept*), tanpa dipengaruhi secara bias oleh probabilitas kelas mayoritas lainnya.

2.3 Teknik Pembagian *Dataset*

Pembagian *dataset* dilakukan untuk memastikan model memiliki data yang cukup untuk belajar sekaligus diuji performanya. Skema pembagian yang digunakan adalah sebagai berikut: 80% Data *training* (5.600 hadis) dan 20% Data *Testing* (1.400 hadis)

2.4 Tahap *Preprocessing* Teks

Setelah melakukan pembagian *dataset*, dilakukan *text preprocessing* yang bertujuan untuk membersihkan dan menyiapkan teks agar berada dalam kondisi yang optimal. Yang kemudian diproses pada tahap selanjujtnya. Adapun tahapannya yaitu sebagai berikut:

a. *Case folding*, yaitu mengubah semua teks menjadi huruf kecil yang ada pada data.
b. *Whitespace Cleaning,* mengurangi spasi berlebih.
c. *String Splitting*, pemisahan tanda baca karena tanda baca tetap diproses menjadi token pada tahap selanjutnya.

Pada penelitian ini, proses stemming dan stopword removal tidak dilakukan. Hal ini disebabkan oleh hasil stemming dan stopword removal cenderung menurunkan akurasi klasifikasi karena dapat menghilangkan struktur asli dari kata yang memiliki makna khusus dalam konteks hadis. [7]

2.5 *Balancing Class*

Pada penelitian ini, penanganan ketidakseimbangan data dilakukan dengan pendekatan yang berbeda untuk tiap jenis label, yaitu mnggunakan penerapan teknik *Data Augmentation* karena karakteristik datanya juga tidak sama. Untuk label anjuran dan larangan, digunakan kombinasi antara back-translation augmentation dan random undersampling. Proses back-translation dilakukan dengan cara menerjemahkan teks dari Bahasa Indonesia ke Bahasa Inggris, kemudian diterjemahkan kembali ke Bahasa Indonesia. Tujuannya supaya jumlah data pada kelas yang lebih sedikit bisa bertambah, tapi tetap mempertahankan makna aslinya. [8]Setelah itu, dilakukan undersampling pada kelas yang jumlahnya lebih banyak agar distribusi data menjadi lebih seimbang. [9]

Sementara itu, untuk label informasi, digunakan penerapan C*ontextual Augmentation* dan H*ybrid Sampling*. [10]Metode ini merupakan gabungan antara oversampling pada kelas minoritas dan undersampling pada kelas mayoritas. Dengan cara ini, jumlah data di kedua kelas bisa disesuaikan ke titik tengah, sehingga tidak terlalu banyak data yang dibuang, tapi juga tidak terlalu banyak duplikasi data seperti pada oversampling murni. [11]

Pemilihan metode yang berbeda ini dilakukan karena setiap label memiliki distribusi data yang berbeda. Teknik augmentasi tidak diterapkan berdasarkan kombinasi label, tetapi berdasarkan instance teks sehingga setiap hadis diperlakukan sebagai satu kesatuan data. Dengan pendekatan tersebut, diharapkan model dapat belajar dengan lebih

ISSN 2774-3659 (Media Online)
Vol 6, No 4, June 2026 | Hal 1666-1674
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i4.1219

seimbang dan tidak cenderung bias ke salah satu kelas saja. Selain itu, teknik ini juga membantu meningkatkan kualitas prediksi, terutama pada kelas yang awalnya jumlah datanya lebih sedikit.

2.6 Ekstraksi Fitur

Adapun langkah selanjutnya yaitu menggunakan model *feature extraction* IndoBERT. [12]Pada tahap ini, teks hadis yang sudah melalui proses preprocessing tidak langsung digunakan untuk klasifikasi, melainkan terlebih dahulu diubah menjadi representasi numerik. Hal ini diperlukan karena model seperti Logistic Regression hanya dapat memproses data dalam bentuk angka.

IndoBERT digunakan sebagai *feature extractor* karena mampu menangkap konteks dan makna dari suatu kalimat dengan lebih baik dibandingkan metode representasi teks tradisional. [13]Setiap teks yang dimasukkan akan diproses oleh tokenizer IndoBERT untuk diubah menjadi token-token yang sesuai dengan kosakata model. Setelah itu, token tersebut akan diproses oleh model untuk menghasilkan representasi vektor.[14]

Dalam penelitian ini, representasi yang diambil adalah dari token khusus [CLS], yaitu token yang berada di awal setiap input. Token ini dianggap mewakili keseluruhan makna dari kalimat. Output dari token [CLS] berupa vektor berdimensi 768, yang kemudian digunakan sebagai fitur input untuk model klasifikasi.[15]

Proses ini dilakukan untuk seluruh data, baik data training, validasi, maupun testing. Namun, model IndoBERT pada tahap ini tidak dilatih ulang (*fine-tuning*), melainkan hanya digunakan sebagai ekstraktor fitur dengan parameter yang sudah tersedia (*pre-trained*). [12]Dengan cara ini, proses menjadi lebih efisien, tetapi tetap mampu menghasilkan representasi teks yang kaya secara semantik.

2.8 Data Validasi

Pada tahapan pelatihan model, *Data training* kembali dibagi menjadi: 80% untuk Data *training* (5.040 hadis) 20% untuk Data validasi (560 hadis). Data validasi digunakan untuk mengevaluasi performa model sebelum diterapkan ke Data *Testing.*

2.9 *Training* Model

Pada tahapan ini, dilakukan dua model latih terhadap dua metode, *baseline* dan optimasi. Pada model *baseline*, metode tidak dilakukan penyetelan parameter sedangkan untuk optimasi beberapa hal dilakukan seperti penyetelan parameter untuk Logistic Regression, melakukan *balancing hybridsampling* serta tambahan optimasi Threshold.

2.10 Logistic Regression

Setelah tahap *feature extraction* menggunakan IndoBERT selesai, langkah selanjutnya adalah melakukan proses klasifikasi menggunakan algoritma Logistic Regression. Logistic Regression merupakan salah satu metode klasifikasi yang cukup sederhana dan banyak digunakan dalam berbagai penelitian, terutama untuk kasus klasifikasi biner.[16] Model ini bekerja dengan memanfaatkan fungsi logistik (sigmoid) untuk mengubah nilai input menjadi probabilitas antara 0 dan 1, yang kemudian digunakan untuk menentukan kelas dari suatu data.

Secara umum, Logistic Regression mencoba mencari hubungan antara fitur input dengan label output melalui proses optimasi.[17] Dalam konteks penelitian ini, fitur input yang digunakan bukan lagi berupa teks mentah, melainkan hasil representasi vektor dari IndoBERT. Dengan demikian, Logistic Regression berperan sebagai model klasifikasi yang memanfaatkan informasi semantik yang telah diekstraksi sebelumnya.[18]

Pada penelitian ini, beberapa parameter penting digunakan untuk mengoptimalkan kinerja model Logistic Regression. Parameter yang digunakan antara lain nilai **C** sebagai pengatur kekuatan regularisasi dengan beberapa variasi nilai seperti 0.01, 0.1, 1, 10, dan 100. Selain itu, digunakan *penalty* berupa L2 karena lebih stabil dalam menangani data berdimensi tinggi seperti hasil embedding.[19] Untuk proses optimasi, digunakan *solver* yaitu *lbfgs* dan *liblinear* yang umum digunakan pada kasus klasifikasi. Parameter *class_weight* juga diuji dengan nilai *None* dan *balanced* untuk melihat pengaruh penyeimbangan bobot kelas terhadap performa model. Proses pemilihan kombinasi parameter terbaik dilakukan menggunakan teknik *Grid Search* dengan *cross-validation* sebanyak 5 lipatan, serta menggunakan metrik evaluasi berupa F1-score.[20]

2.11 Evaluasi Model

Setelah model Logistic Regression dilatih menggunakan fitur hasil ekstraksi IndoBERT, tahap selanjutnya adalah melakukan evaluasi untuk melihat seberapa baik model dalam melakukan klasifikasi. Evaluasi ini penting karena dari sini bisa diketahui apakah model sudah mampu membedakan setiap kelas dengan baik atau masih terdapat kesalahan yang cukup signifikan.[21]

Evaluasi model hanya berfokus pada metrik, *F1-score*. F1-score merupakan gabungan dari precision dan recall yang memberikan gambaran keseimbangan antara keduanya. Metrik ini dipilih karena dataset yang digunakan memiliki distribusi yang tidak seimbang, sehingga akurasi saja tidak cukup untuk menggambarkan performa model secara keseluruhan.

Selain itu, model juga dievaluasi menggunakan *Average Precision (AP)* yang diperoleh dari kurva *Precision-Recall*. Kurva ini digunakan untuk melihat hubungan antara precision dan recall pada berbagai nilai threshold. Dari kurva

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 4, June 2026 | Hal 1666-1674
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i4.1219

tersebut, dapat diketahui bagaimana perubahan threshold mempengaruhi performa model, terutama dalam menentukan batas keputusan antara kelas positif dan negatif.

Dalam penelitian ini, nilai threshold tidak langsung menggunakan nilai default sebesar 0.5, melainkan dicari nilai yang paling optimal berdasarkan data validasi. Beberapa nilai threshold diuji secara manual, kemudian dipilih nilai yang menghasilkan F1-score tertinggi. Pendekatan ini dilakukan karena pada beberapa kasus, terutama ketika data tidak seimbang, penggunaan threshold default seringkali tidak memberikan hasil terbaik. [22]Rumus evaluasi ditunjukkan pada persamaan berikut ini [25].

Dengan:

$$Precision = \frac{TP}{TP+FP} \quad (1)$$

$$Recall = \frac{TP}{TP+FN} \quad (2)$$

Pada persamaan diatas Presisi dihitung sebagai rasio antara jumlah prediksi positif yang benar (*True Positive*) terhadap total seluruh prediksi positif yang dilakukan oleh model, baik yang benar maupun yang salah (*True Positive* ditambah *False Positive*). Sementara itu, *Recall* dihitung sebagai rasio antara jumlah prediksi positif yang benar (*True Positive*) terhadap total seluruh data yang seharusnya berlabel positif (*True Positive* ditambah *False Negative*).

$$AP = \sum_n (R_n - R_{n-1}) \times P_n \quad (3)$$

Kemudian *Average Precision* (AP) adalah metrik yang digunakan untuk menghitung rata-rata nilai precision pada setiap titik perubahan recall *Di mana:*P_n (*Precision* pada titik ke-n) Merupakan nilai presisi model pada titik evaluasi ke-n. R_n (*Recall* pada titik ke-n) Merupakan nilai *recall* model pada titik evaluasi ke-n. sedangkan $(R_n - R_{n-1})$ Merupakan selisih atau perubahan nilai *recall* di antara dua titik evaluasi yang berurutan.

Setelah threshold optimal diperoleh, nilai tersebut kemudian digunakan untuk melakukan prediksi akhir pada data validasi dan data uji. Hasil evaluasi menunjukkan bahwa penyesuaian threshold dapat meningkatkan performa model, terutama dalam hal keseimbangan antara precision dan recall. Meskipun demikian, masih terdapat beberapa kesalahan klasifikasi yang menunjukkan bahwa model belum sepenuhnya optimal dalam memahami konteks teks hadis.

2.12 Analisis Eror

Dalam melakukan penelitian, masih ditemukan beberapa kesalahan klasifikasi yang memengaruhi performa model. Salah satu penyebab utamanya adalah model masih mengalami kesulitan dalam mengklasifikasikan kelas minoritas meskipun telah dilakukan balancing dan augmentasi data. Kondisi ini disebabkan oleh beberapa faktor, yaitu distribusi data validasi dan pengujian yang tetap tidak seimbang, kualitas hasil contextual augmentation yang tidak selalu mempertahankan makna asli sehingga dapat menimbulkan noise, serta kemiripan struktur teks hadis yang membuat perbedaan antar kelas menjadi kurang jelas. Selain itu, adanya batas maksimum token pada proses embedding menyebabkan sebagian informasi pada hadis yang panjang tidak seluruhnya diproses oleh model. Perbedaan performa yang cukup besar antara data pelatihan dan data validasi maupun pengujian juga mengindikasikan adanya kecenderungan overfitting, sehingga kemampuan generalisasi model terhadap data baru masih terbatas.

3. HASIL DAN PEMBAHASAN

Penelitian ini menggunakan bahasa pemrograman Python untuk mengolah data terjemahan Hadis Bukhari. Proses penelitian dilakukan melalui dua skenario, yaitu skenario awal dan skenario optimal. Pada skenario awal (baseline), klasifikasi dilakukan tanpa penerapan teknik balancing dan masih menggunakan parameter bawaan sistem. Adapun pada skenario optimal, proses klasifikasi dikembangkan dengan menerapkan teknik balancing serta melakukan optimasi parameter melalui hyperparameter tuning guna meningkatkan performa hasil klasifikasi.

3.1 Model *Baseline*

Pengujian pertama dilakukan tanpa menggunakan teknik balancing, tanpa *tuning threshold* dan menggunakan parameter *default* pada masing-masing label.

Tabel 2. Hasil Model Awal Metode LR pada Data Latih dan Data Validasi

Kategori	Data Latih		Data Validasi	
	Accuracy	f1-score	accuracy	f1-score
Anjuran	87,34%	76,29%	83,04%	68,13%
Larangan	94,20%	84,66%	89,82%	73,20%
Informasi	96,58%	77,19%	94,02%	58,09%

Berdasarkan Tabel 2, pengujian awal menggunakan metode *Logistic Regression* pada data latih dan validasi menunjukkan performa yang cukup baik pada seluruh kategori. Kategori informasi memperoleh nilai *accuracy* tertinggi, yaitu 96,58% pada data latih dan 94,02% pada data validasi, namun nilai *F1-Score* masih lebih rendah dibanding kategori

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 4, June 2026 | Hal 1666-1674
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i4.1219

lain. Kategori larangan menunjukkan performa paling stabil dengan *F1-Score* validasi tertinggi sebesar 73,20%. Sedangkan kategori anjuran memiliki performa paling rendah. Untuk melihat kemampuan generalisasi model, pengujian juga dilakukan pada data *test* dan hasilnya disajikan pada Tabel 3.

Tabel 3. Hasil Model Awal Metode LR pada Data Uji

Kategori	Data Test	
	accuracy	f1-score
anjuran	82,43%	65,81%
larangan	90,21%	72,38%
informasi	94,21%	58,40%

Berdasarkan hasil pada Tabel 3, metode Logistic Regression mampu menghasilkan tingkat akurasi yang tinggi pada data uji untuk seluruh kategori, dengan nilai berkisar antara 82% hingga 94% serta rata-rata mencapai 88,95%. Meskipun demikian, nilai F1-score pada kategori anjuran dan informasi masih tergolong rendah dibanding kategori lainnya.

3.2 Model Optimasi

Setelah melakukan *baseline,* penelitian dilanjutkan dengan melakukan optimasi terbaik yaitu dengan menggunakan teknik *data augmentation* berbasis *back translation* pada label Anjuran dan Larangan, serta *contextual augmentation* pada label Informasi. Perbedaan teknik augmentasi dilakukan dengan mempertimbangkan karakteristik semantik setiap kategori hadis. Label Anjuran dan Larangan memiliki ketergantungan yang tinggi terhadap makna normatif yang terkandung dalam teks, sehingga digunakan metode Back Translation untuk menghasilkan variasi kalimat dengan risiko perubahan makna yang lebih rendah. Sebaliknya, label Informasi cenderung berisi narasi dan deskripsi yang lebih toleran terhadap variasi leksikal, sehingga Contextual Augmentation dipilih karena mampu menghasilkan keragaman kosakata berdasarkan konteks kalimat. Pendekatan ini dapat meningkatkan jumlah dan variasi data tanpa menghilangkan karakteristik utama dari masing-masing kategori.[23] Kemudian Optimasi *Threshold,* dan *Hyperparameter* menggunakan *RandomizedSearhCV. RandomizedSearhCV* digunakan karena teknik *tuning* yang memiliki waktu yang singkat dalam proses optimasi. [24] Lalu pengunaan *Threshold tuning* dilakukan pada data validasi untuk menentukan nilai ambang optimal klasifikasi. Nilai threshold ditentukan dengan memaksimalkan skor F1 pada kurva precision–recall. Pendekatan ini dipilih karena dataset hadis memiliki distribusi kelas yang tidak seimbang. Threshold optimal yang diperoleh kemudian diterapkan pada data uji untuk menghasilkan prediksi akhir.Tabel 4 merupakan hasil dari model optimal pada metode Logistic Regression.

Tabel 4. Hasil Model Optimasi Metode LR pada Data Latih dan Data Validasi

Kategori	Data Latih		Data Validasi	
	Accuracy	f1-score	Accuracy	f1-score
Anjuran	90,11%	90,11%	80,98%	70,33%
Larangan	97,22%	97,22%	87,59%	73,77%
Informasi	96,57%	96,57%	93,13%	61,53%

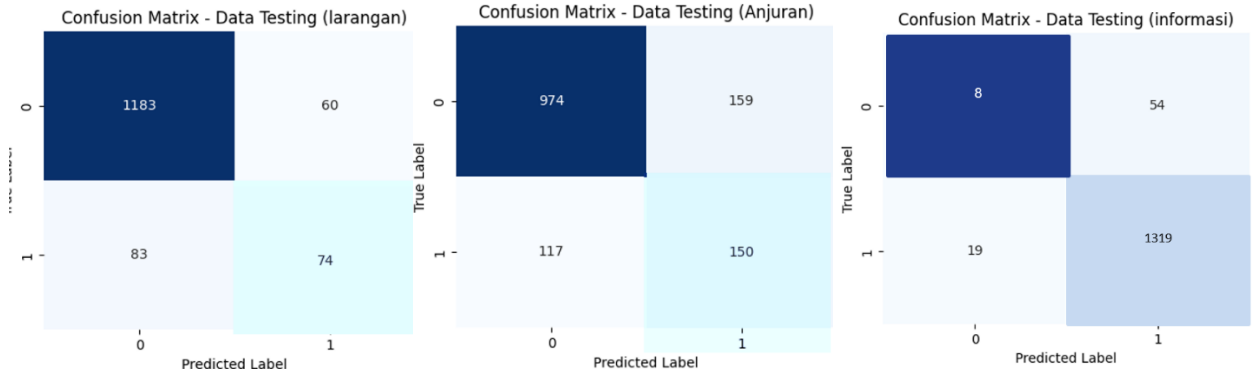
Berdasarkan Tabel 4 hasil pengujian optimal menunjukkan data *train* kategori Larangan memperoleh performa terbaik yaitu sebesar 97,22%, kategori Informasi sebesar 96,57%, sedangkan kategori *Anjuran* sebesar 90,11%. Sementara itu, pada data validasi diperoleh nilai *accuracy* dan *f1-score* sebesar 80,98% dan 70,33% untuk kategori *Anjuran*, 87,59% dan 73,77% untuk kategori *Larangan*, serta 93,13% dan 61,53% untuk kategori *Informasi*. Meskipun terjadi penurunan performa dari data *train* ke data validasi, selisih nilai yang dihasilkan masih tergolong wajar sehingga menunjukkan bahwa model memiliki kemampuan generalisasi yang cukup baik dan tidak mengalami *overfitting* secara signifikan. Selanjutnya Untuk mengevaluasi lebih lanjut generalisasi model, metode Logistic Regression (LR) juga diuji menggunakan data uji (test). Hasil pengujian tersebut disajikan pada Tabel 5.

Tabel 5. Hasil Model Optimal Metode LR pada Data Uji

Kategori	Data Test	
	accuracy	f1-score
Anjuran	80,29%	69,84%
larangan	88,14%	72,35%
informasi	93,36%	60,46%

Hasil optimal pada metode RF pada data uji menunjukkan performa yang bervariasi antar kategori. Kategori *Larangan* memperoleh hasil terbaik dengan nilai *accuracy* sebesar 88,14% dan *f1-score* sebesar 72,35%. Selanjutnya, kategori *Anjuran* menghasilkan *accuracy* sebesar 80,29% dengan *f1-score* 69,84%. Adapun kategori *Informasi* memperoleh *accuracy* tertinggi sebesar 93,36%, namun nilai *f1-score* yang dihasilkan sebesar 60,46%. Hal ini menunjukkan bahwa meskipun akurasi pada semua kategori tergolong tinggi, nilai *F1-Score*, masih menjadi patokan utama, terutama pada kategori informasi, masih perlu ditingkatkan. Berikut hasil Confusion Matriks pada masing-masing kategori.

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 4, June 2026 | Hal 1666-1674
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i4.1219



Gambar 2. *Confusion Matrix* pada kategori Anjuran, Larangan, dan Informasi untuk data Testing Model Optimal

Berdasarkan Gambar 2 confusion matrix pada kategori Anjuran untuk data testing, model berhasil mengklasifikasikan 974 data yang tidak mengandung unsur anjuran (kelas 0) dengan benar (true negative) dan 150 data yang mengandung unsur anjuran (kelas 1) dengan benar (true positive). Di sisi lain, terdapat 159 data kelas 0 yang salah diprediksi sebagai kelas 1 (false positive) serta 117 data kelas 1 yang salah diprediksi sebagai kelas 0 (false negative).

Hasil tersebut menunjukkan bahwa model memiliki kemampuan yang baik dalam mengenali kelas mayoritas, yaitu kelas 0, yang ditunjukkan oleh jumlah true negative yang relatif tinggi. Selain itu, jumlah false negative yang lebih rendah dibandingkan true positive mengindikasikan bahwa sebagian besar hadis yang mengandung unsur anjuran berhasil diidentifikasi dengan benar oleh model. Meskipun masih terdapat kesalahan klasifikasi pada kedua kelas, distribusi nilai pada confusion matrix menunjukkan bahwa model mampu membedakan hadis yang mengandung dan tidak mengandung unsur anjuran dengan cukup baik pada data testing, yang sejalan dengan nilai F1-score kategori Anjuran sebesar 69,84%.

Kemudian *Confusion Matrix* pada kategori Larangan, model berhasil mengklasifikasikan 1.183 data yang tidak mengandung unsur larangan (kelas 0) dengan benar (true negative) dan 74 data yang mengandung unsur larangan (kelas 1) dengan benar (true positive). Sementara itu, terdapat 60 data kelas 0 yang salah diprediksi sebagai kelas 1 (false positive) dan 83 data kelas 1 yang salah diprediksi sebagai kelas 0 (false negative).

Hasil tersebut menunjukkan bahwa model memiliki kemampuan yang baik dalam membedakan data yang mengandung dan tidak mengandung unsur larangan. Jumlah true negative yang tinggi mengindikasikan bahwa sebagian besar hadis yang tidak termasuk kategori larangan dapat dikenali dengan benar oleh model. Selain itu, jumlah false positive yang relatif rendah menunjukkan bahwa model tidak banyak melakukan kesalahan dalam mengklasifikasikan hadis non-larangan sebagai larangan.

Meskipun masih terdapat false negative sebanyak 83 data, yang menunjukkan bahwa beberapa hadis berkategori larangan belum berhasil teridentifikasi, secara keseluruhan distribusi nilai pada confusion matrix memperlihatkan bahwa model memiliki kemampuan klasifikasi yang baik dan cukup seimbang pada data testing. Hal ini sejalan dengan nilai F1-score kategori Larangan sebesar 72,35%, yang merupakan salah satu performa terbaik di antara kategori yang diuji, sehingga menunjukkan bahwa model mampu mengidentifikasi pola pada kategori larangan dengan cukup efektif.

Selanjutnya *Confusion Matrix* pada kategori Informasi untuk data Testing, model berhasil mengklasifikasikan 1.319 data yang mengandung unsur informasi (kelas 1) sebagai true positive dan 8 data yang tidak mengandung unsur informasi (kelas 0) sebagai true negative. Sementara itu, terdapat 54 data kelas 0 yang salah diprediksi sebagai kelas 1 (false positive) dan 19 data kelas 1 yang salah diprediksi sebagai kelas 0 (false negative). Hasil ini menunjukkan bahwa model lebih banyak menghasilkan prediksi yang benar dibandingkan prediksi yang salah pada kategori informasi.

Hal ini menunjukkan model memiliki kemampuan yang sangat baik dalam mengidentifikasi hadis yang mengandung unsur informasi, yang ditunjukkan oleh tingginya jumlah true positive dan rendahnya jumlah false negative. Namun, jumlah true negative yang relatif sedikit dibandingkan false positive mengindikasikan bahwa model cenderung memprediksi data sebagai kategori informasi. Kondisi ini sejalan dengan karakteristik distribusi data yang didominasi oleh kelas informasi, sehingga meskipun model memperoleh nilai accuracy yang tinggi, nilai F1-score sebesar 60,46% menunjukkan bahwa keseimbangan antara precision dan recall pada kategori ini masih dapat ditingkatkan.

3.3 Pembahasan

Penelitian ini melakukan perbandingan terhadap beberapa penelitian terdahulu yang membahas klasifikasi teks menggunakan berbagai metode machine learning. Beberapa penelitian sebelumnya memanfaatkan algoritma Support Vector Machine, Long Short-Term Memory, serta Random Forest untuk mengidentifikasi kategori hadis berdasarkan isi terjemahannya. Penelitian lain juga menggunakan TF-IDF sebagai feature extraction yang dikombinasikan dengan Logistic Regression dalam proses klasifikasi teks. Perbandingan dilakukan berdasarkan metode yang digunakan, jenis feature extraction, dataset, serta nilai evaluasi model yang diperoleh. Tujuan dari perbandingan ini adalah untuk mengetahui efektivitas penggunaan IndoBERT sebagai feature extraction dan Logistic Regression sebagai classifier dalam klasifikasi multilabel pada terjemahan hadis Bukhari. Ringkasan perbandingan penelitian sebelumnya dapat dilihat pada Tabel 6.

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 4, June 2026 | Hal 1666-1674
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i4.1219

Tabel 6. Perbandingan dengan penelitian terdahulu

Metode	Feature Extraction	F1-Score Anjuran	F1-Score Larangan	F1-Score Informasi
SVM	TF-IDF	69.41	82.57	67.40
RF	TF-IDF	69.03	84.45	58.14
LSTM	Embedding Layer	71.04	84.04	64.27
Logistic Regression	TF-IDF	73.95	79.86	57.23
Logistic Regression	IndoBERT	69.84	72.35	60.11

Berdasarkan hasil perbandingan dengan penelitian terdahulu, model Logistic Regression berbasis ekstraksi fitur IndoBERT menunjukkan performa yang kompetitif pada klasifikasi multilabel terjemahan Hadis Bukhari. Pada kategori Anjuran, model yang diusulkan memperoleh F1-*score* sebesar 69,84%, dapat mengungguli dari metode SVM dan RF. Namun sedikit lebih rendah dibandingkan metode Logistic Regression berbasis TF-IDF yang mencapai 73,95% serta metode LSTM sebesar 71,04%.

Pada kategori Larangan, metode yang diusulkan menghasilkan F1-*score* sebesar 72,35%. Nilai tersebut masih berada di bawah metode berbasis TF-IDF dan LSTM, di mana Random Forest memperoleh performa tertinggi sebesar 84,45%. Perbedaan hasil ini menunjukkan bahwa representasi fitur berbasis IndoBERT belum sepenuhnya optimal untuk label Larangan, meskipun telah dilakukan optimasi model.

Sementara itu, pada kategori Informasi, model Logistic Regression dengan fitur IndoBERT memperoleh F1-*score* sebesar 60,11%, lebih baik dibandingkan Logistic Regression berbasis TF-IDF yang hanya mencapai 57,23% dan Random Forest sebesar 58,14%. Hasil ini menunjukkan bahwa representasi kontekstual dari IndoBERT mampu memberikan peningkatan performa pada label yang memiliki variasi konteks kalimat lebih kompleks.

4. KESIMPULAN

Penelitian ini berhasil mensintesis sebuah kerangka kerja klasifikasi teks multilabel pada terjemahan Hadis Bukhari yang efisien secara komputasi melalui kombinasi ekstraksi fitur IndoBERT dan Logistic Regression. Temuan penelitian menunjukkan bahwa representasi semantik beku (frozen embeddings) dari IndoBERT memiliki kapasitas yang kokoh dalam memetakan batas-batas hukum fikih yang kompleks (Anjuran, Larangan, dan Informasi). Kontribusi praktis dari penelitian ini terletak pada pembuktian bahwa pendekatan arsitektur yang ringan (lightweight architecture) ini, jika dipadukan dengan teknik augmentasi data bertarget (back-translation dan contextual augmentation) serta optimasi threshold, mampu menghasilkan generalisasi prediksi yang stabil dan seimbang antar-kelas, tanpa memerlukan sumber daya komputasi yang besar. Meskipun terdapat trade-off marjinal pada nilai akurasi jika dibandingkan dengan model deep learning end-to-end yang padat parameter, model yang diusulkan menawarkan keunggulan nyata dari segi efisiensi memori, kecepatan inferensi, dan kemudahan deployment pada lingkungan dengan komputasi terbatas (seperti aplikasi mobile atau API server skala kecil). Dengan demikian, penelitian ini memberikan kontribusi teoritis bahwa rekayasa pada tingkat data (augmentasi) dan penyesuaian ambang batas keputusan (threshold tuning) secara signifikan dapat mengompensasi keterbatasan model linier dalam menangani ruang fitur berdimensi tinggi (768 dimensi) pada teks keagamaan klasik. Sebagai arah pengembangan ke depan, fokus penelitian sebaiknya tidak diarahkan pada memperbesar skala arsitektur model, melainkan pada pengayaan semantik di tingkat hulu. Eksplorasi berikutnya dapat diarahkan pada integrasi ontologi hukum Islam (Fikih) ke dalam ruang vektor embedding untuk memperkuat pemahaman konteks hadis yang bersifat implisit. Selain itu, pengujian ketangguhan generalisasi model perlu diperluas menggunakan pendekatan cross-validation lintas kitab hadis primer lainnya guna menguji konsistensi performa model dalam domain teks klasik yang lebih luas.

REFERENCES

[1] M. A. H. Muhammad, M. I. Afkarina, S. S. Shalsabila, and S. Fikri, "Hadist Ditinjau Dari Kualitas Sanad Dan Matan (Hadist Shohih, Hasan, Dhoif)," *Jurnal Kajian Islam dan Sosial Keagamaan*, vol. 1, no. 4, pp. 396–401, 2024, [Online]. Available: https://jurnal.ittc.web.id/index.php/jkis/article/view/1103

[2] R. S. Mutaqin, Z. Nurpadilah, and H. Z. Muttaqin, "Perawi Mudallis dalam Shahih Bukhari: Studi al-Jarh wa al-Ta'dil pada 'Umar bin 'Ali bin 'Atha' bin Muqaddam," *Riwayah: Jurnal Studi Hadis*, vol. 7, no. 2, pp. 241–272, Dec. 2022, doi: 10.21043/riwayah.v7i2.10651.

[3] Y. Sagama and A. Alamsyah, "Multi-label classification of Indonesian online toxicity using BERT and RoBERTa," in *2023 IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology (IAICT)*, 2023, pp. 143–149.

[4] A. Ramadhani, N. S. Harahap, S. Agustian, I. Iskandar, and S. Sanjaya, "Perbandingan Performa Metode Klasifikasi Teks Multi-Label Terjemahan Hadis Bukhari Menggunakan Metode Support Vector Machine Dan Long Short Term Memory," *MALCOM*, 2024, doi: 10.57152/malcom.v5i3.2051.

[5] R. Z. N. Ahmad, N. S. Harahap, S. Agustian, I. Iskandar, and S. Sanjaya, "Perbandingan Performa Random Forest dan Long Short-Term Memory dalam Klasifikasi Teks Multilabel Terjemahan Hadits Bukhari," *MALCOM*, 2024, doi: 10.57152/malcom.v5i3.2046.

[6] G. A. Ghifari, "Klasifikasi Multi Label Terjemahan Hadits Sahih Bukhari Menggunakan Metode Logistic Regression," Politeknik Negeri Jember, 2025.

[7] R. Kustiawan, A. Adiwijaya, and M. D. Purbolaksono, "A Multi-label Classification on Topic of Hadith Verses in Indonesian Translation using CART and Bagging," *Jurnal Media Informatika Budidarma*, vol. 6, no. 2, pp. 868–876, 2022, doi: 10.30865/mib.v6i2.3787.

[8] D. Držík and L. Kelebercová, "Back-translation effects on static and contextual word embeddings for topic classification embedding in classification tasks," *PLoS One*, vol. 20, no. 8, p. e0330622, 2025.
[9] S. Ningsih, N. H. Safaat, S. Agustian, Yusra, and E. P. Cynthia, "Pengaruh Penyeimbangan Data Pada Klasifikasi Terjemahan Al-Quran Dengan Metode Naïve Bayes dan Long Short Term Memory," *Journal of Computer System and Informatics (JoSYC)*, vol. 5, no. 3, pp. 626–635, 2024, doi: 10.47065/josyc.v5i3.5181.
[10] S. Kobayashi, "Contextual Augmentation: Data Augmentation by Words with Paradigmatic Relations," *CoRR*, vol. abs/1805.06201, 2021, [Online]. Available: http://arxiv.org/abs/1805.06201
[11] F. Taufiqurrahman, S. Al Faraby, and M. D. Purbolaksono, "Klasifikasi teks multi label pada hadis terjemahan bahasa indonesia menggunakan chi-square dan svm," *eProceedings of Engineering*, vol. 8, no. 5, 2021, [Online]. Available: https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/15671/15384
[12] Y. Santoso and S. Candra, "Comparative Sentiment Analysis of App Reviews Using TF-IDF and IndoBERT as Feature Extraction with SVM," in *2025 5th International Conference on Intelligent Cybernetics Technology & Applications (ICICyTA)*, 2025, pp. 13–18. DOI:10.1109/ICICyTA68677.2025.11362384.
[13] Y. Asri, D. Kuswardani, W. N. Suliyanti, Y. O. Manullang, and A. R. Ansyari, "Sentiment analysis based on Indonesian language lexicon and IndoBERT on user reviews PLN mobile application," *Indonesian Journal of Electrical Engineering and Computer Science*, 2025, [Online]. Available: https://api.semanticscholar.org/CorpusID:275923734
[14] E. Issa and A. A. Ibrahim, "Integrating BERT for Nuanced Sentiment Analysis: A Detailed Examination of Diverse Textual Datasets," *IEEE Access*, vol. 12, pp. 186296–186312, 2024, [Online]. Available: https://api.semanticscholar.org/CorpusID:274823691
[15] Z. F. Ramadhan and A. B. Mutiara, "Sentiment Analysis of Honkai: Star Rail Indonesian Language Reviews on Google Play Store Using Bidirectional Encoder Representations from Transformers Method," *International Journal of Engineering, Science and Information Technology*, 2023, [Online]. Available: https://api.semanticscholar.org/CorpusID:262177905
[16] D. Dey *et al.*, "The proper application of logistic regression model in complex survey data: a systematic review," *BMC Med. Res. Methodol.*, vol. 25, 2025, [Online]. Available: https://api.semanticscholar.org/CorpusID:275816747
[17] S. Guan, "Clasification Medical Claim Denial Using Logistic Regression and Decision Tree Algorithm," *2024 3rd International Conference on Health Big Data and Intelligent Healthcare (ICHIH)*, pp. 7–10, 2024, [Online]. Available: https://api.semanticscholar.org/CorpusID:279987477
[18] A. Ramzan, R. H. Ali, N. Ali, and A. Khan, "Enhancing Fake News Detection Using BERT: A Comparative Analysis of Logistic Regression, RFC, LSTM and BERT," *2024 International Conference on IT and Industrial Technologies (ICIT)*, pp. 1–6, 2024, [Online]. Available: https://api.semanticscholar.org/CorpusID:276118442
[19] K. Josephine *et al.*, "Comprehensive review of logistic regression techniques in predicting health outcomes and trends," *World Journal of Advanced Pharmaceutical and Life Sciences*, 2024, [Online]. Available: https://api.semanticscholar.org/CorpusID:274507228
[20] Y. Miyazaki *et al.*, "Logistic regression analysis and machine learning for predicting post-stroke gait independence: a retrospective study," *Sci. Rep.*, vol. 14, 2024, [Online]. Available: https://api.semanticscholar.org/CorpusID:272594675
[21] H. Wu, B. Liao, T. Ji, K. Ma, Y. Luo, and S. Zhang, "Comparison between traditional logistic regression and machine learning for predicting mortality in adult sepsis patients," *Front. Med. (Lausanne).*, vol. 11, 2025, [Online]. Available: https://api.semanticscholar.org/CorpusID:275377173
[22] S. Guan, "Clasification Medical Claim Denial Using Logistic Regression and Decision Tree Algorithm," *2024 3rd International Conference on Health Big Data and Intelligent Healthcare (ICHIH)*, pp. 7–10, 2024, [Online]. Available: https://api.semanticscholar.org/CorpusID:279987477
[23] H. Fauzan, A. Adiwijaya, and S. Al-Faraby, "Pengklasifikasian Topik Hadits Terjemahan Bahasa Indonesia Menggunakan Latent Semantic Indexing dan Support Vector Machine," *Jurnal Media Informatika Budidarma*, vol. 2, no. 4, p. 131, 2022.
[24] S. Diantika, H. Nalatissifa, N. Maulidah, R. Supriyadi, and A. Fauzi, "Penerapan Teknik Random Oversampling Untuk Memprediksi Ketepatan Waktu Lulus Menggunakan Algoritma Random Forest," *Computer Science (CO-SCIENCE*, vol. 4, no. 1, pp. 11–18, 2024, doi: 10.31294/coscience.v4i1.1996.