Infotek : Jurnal Informatika dan Teknologi
Vol. 9 No. 2, Juli 2026
Hal.651-662

e-ISSN 2614-8773

DOI : 10.29408/jit.v9i2.35011 Link : https://dx.doi.org/10.29408/jit.v9i2.35011

Perbandingan SVM dan Logistic Regression untuk Klasifikasi Teks Konsultasi Daring Kehamilan dan Menstruasi

Zahra Syifa Prasasti[1], Safitri Juanita[2*]
[1,2] Sistem Informasi, Universitas Budi Luhur
*safitri.juanita@budiluhur.ac.id

Abstrak

Tingginya volume teks jawaban dokter pada konsultasi kesehatan daring, khususnya pada topik kehamilan dan menstruasi, mendorong kebutuhan akan sistem klasifikasi teks medis yang otomatis dan akurat. Namun, kedua topik tersebut sering memiliki kemiripan istilah medis dan konteks konsultasi, sehingga menyulitkan proses klasifikasi. Penelitian ini bertujuan untuk membangun model klasifikasi otomatis pada teks jawaban dokter pada konsultasi kesehatan daring berbahasa Indonesia dengan membandingkan dua algoritma, yaitu Support Vector Machine (SVM) dan Logistic Regression (LR) yang dipilih karena banyak digunakan pada tugas klasifikasi teks dengan TF-IDF serta mampu menangani data berdimensi tinggi. Kontribusi penelitian ini adalah pengembangan model klasifikasi pada dataset jawaban dokter berbahasa Indonesia yang masih jarang diteliti, analisis feature importance untuk mengidentifikasi istilah medis dominan pada setiap kategori, serta evaluasi perbandingan performa SVM dan LR yang diverifikasi menggunakan uji statistik McNemar. Data penelitian menggunakan dataset publik "Doctor's Answer Text Dataset in Indonesian" dari Mendeley Data yang terdiri atas 17.807 data. Model dibangun menggunakan pendekatan CRISP-DM dengan ekstraksi fitur TF-IDF dan dievaluasi pada empat rasio pembagian data menggunakan metrik akurasi, presisi, recall, dan F1-score. Hasil menunjukkan bahwa LR secara konsisten mengungguli SVM dengan akurasi terbaik 0.9359 pada rasio 80:20. Analisis feature importance mengidentifikasi kata "hamil" dan "janin" sebagai fitur dominan pada kategori kehamilan, serta "menstruasi" dan "siklus" pada kategori menstruasi. Temuan ini menunjukkan bahwa LR berbasis TF-IDF efektif untuk klasifikasi teks konsultasi kesehatan daring berbahasa Indonesia.

.

Kata kunci : feature importance, kesehatan reproduksi, klasifikasi teks medis, logistic regression, SVM.

Abstract

The high volume of doctors' response texts in online health consultations, particularly on pregnancy and menstruation topics, has increased the need for an accurate and automated medical text classification system. However, these two topics often share similar medical terminology and consultation contexts, making the classification process challenging. This study aims to develop an automatic classification model for Indonesian-language doctors' responses in online health consultations by comparing two algorithms, namely Support Vector Machine (SVM) and Logistic Regression (LR), which were selected because they are widely used in text classification tasks with TF-IDF feature representation and are capable of handling high-dimensional data. The contributions of this study include the development of a classification model using an underexplored Indonesian doctors' response dataset, feature importance analysis to identify dominant medical terms in each category, and a comparative evaluation of SVM and LR performance validated using McNemar's statistical test. The study utilized the public "Doctor's Answer Text Dataset in Indonesian" from Mendeley Data, comprising 17,807 records. The models were developed using the CRISP-DM framework with TF-IDF feature extraction and evaluated across four train-test split ratios using accuracy, precision, recall, and F1-score metrics. The results showed that Logistic Regression (LR) consistently outperformed Support Vector Machine (SVM), achieving the highest accuracy of 0.9359 with an 80:20 train-test split. Feature importance analysis identified "hamil" (pregnant) and "janin" (fetus) as the dominant features for the pregnancy category, and "menstruasi" (menstruation) and "siklus" (cycle) for the menstruation category. These findings demonstrate that TF-IDF-based LR is an effective approach for classifying Indonesian-language online health consultation texts.

Keywords : *feature importance, logistic regression, medical text classification, reproductive health, SVM.*

Infotek : Jurnal Informatika dan Teknologi
Vol. 9 No. 2, Juli 2026
Hal.651-662

e-ISSN 2614-8773

DOI : 10.29408/jit.v9i2.35011 Link : https://dx.doi.org/10.29408/jit.v9i2.35011

1. Pendahuluan

Perkembangan teknologi digital mendorong telemedis menjadi bagian penting layanan kesehatan di Indonesia. Platform seperti Halodoc, Alodokter, dan KlikDokter menyediakan konsultasi daring, pengiriman obat, serta layanan berbasis kecerdasan buatan bagi jutaan pengguna [1]. Namun, pertumbuhan yang pesat ini membawa tantangan berupa pengelolaan volume data yang besar, keterbatasan literasi digital, serta dukungan regulasi yang belum komprehensif [2]. Di antara seluruh layanan yang tersedia, kesehatan reproduksi wanita khususnya mengenai kehamilan dan menstruasi menjadi topik dengan frekuensi pencarian informasi dan konsultasi daring tertinggi di Indonesia. Pada topik kehamilan, penggunaan aplikasi kesehatan *mobile* di kalangan ibu hamil usia 20–35 tahun terus meningkat untuk memperoleh informasi dan edukasi digital terkait kehamilan [3].

Kondisi serupa juga terjadi pada topik menstruasi, di mana penggunaan aplikasi pelacak siklus menstruasi sebagai bagian dari inovasi *mobile health* terus meningkat untuk mendukung kesehatan reproduksi perempuan [4]. Tingginya kebutuhan konsultasi pada kedua topik ini didorong oleh besarnya prevalensi gangguan yang menyertainya, data menunjukkan bahwa 90% remaja Gen Z mengalami gangguan siklus menstruasi [5], sementara tingginya angka kematian ibu mendorong wanita untuk lebih aktif mencari informasi kehamilan secara daring untuk mendeteksi risiko sejak dini [6].

Selain memiliki volume konsultasi yang tinggi, topik kehamilan dan menstruasi juga mengandung beragam istilah medis, gejala, dan konteks klinis yang saling berkaitan tetapi memerlukan penanganan dan interpretasi medis yang berbeda. Karakteristik tersebut menjadikan kedua topik ini penting untuk dianalisis dalam penelitian klasifikasi teks kesehatan. Akibatnya, *platform* konsultasi kesehatan daring menghasilkan data teks dalam jumlah besar dan beragam. Penelitian ini berfokus pada teks jawaban dokter karena merepresentasikan pengetahuan klinis, dengan terminologi klinis yang lebih konsisten, terstruktur, dan informasi medis yang komprehensif dibandingkan teks keluhan pasien yang menggunakan istilah sehari-hari sehingga lebih sulit dianalisis secara klinis.

Karakteristik data tersebut menunjukkan perlunya pendekatan otomatis berbasis *machine learning* untuk mengklasifikasikan topik konsultasi daring dari data teks yang besar dan tidak terstruktur, dengan ekstraksi fitur TF-IDF karena mampu merepresentasikan istilah penting dalam teks kesehatan dan membantu meningkatkan performa algoritma klasifikasi.

2. Tinjauan Pustaka

2.1. Penelitian Terkait

Beberapa penelitian terdahulu terkait klasifikasi

Infotek : Jurnal Informatika dan Teknologi
Vol. 9 No. 2, Juli 2026
Hal.651-662

e-ISSN 2614-8773

DOI : 10.29408/jit.v9i2.35011 Link : https://dx.doi.org/10.29408/jit.v9i2.35011

teks dengan data medis menggunakan SVM dan Logistic Regression (LR) antara lain:

1. Penelitian klasifikasi data medis penyakit Diabetes Mellitus berbasis teks menggunakan TF-IDF, membuktikan bahwa SVM efektif, dengan akurasi tertinggi mencapai 98.83% [7].

2. Kemudian penelitian lain menggunakan data teks kesehatan mental menerapkan algoritma Logistic Regression berbasis TF-IDF dan teknik SMOTE pada 53.042 data menunjukkan bahwa model tanpa SMOTE menghasilkan akurasi lebih tinggi sebesar 76.9%, yang mengindikasikan bahwa SMOTE tidak selalu meningkatkan performa klasifikasi teks [8]. Penelitian ini memiliki keterkaitan metodologis karena kesamaan penggunaan LR dan TF-IDF, namun pada karakteristik teks berbeda.

3. Penelitian lainnya menggunakan dataset rekam medis klinis dan dataset literatur medis, dengan pendekatan *hybrid machine learning* yang menggabungkan algoritma tradisional dan *Adaptive Genetic Algorithm* membuktikan bahwa pendekatan *hybrid* dengan optimasi parameter terbukti meningkatkan performa dibandingkan algoritma tunggal [9].

4. Sementara itu, penelitian lain membandingkan Naïve Bayes dan Random Forest pada 27.977 data teks kesehatan mental berbasis TF-IDF. Penelitian ini menemukan bahwa Random Forest unggul dengan akurasi 89.3%, sehingga membuktikan bahwa klasifikasi teks berbasis TF-IDF pada domain kesehatan dapat menghasilkan performa yang tinggi [10]. Penelitian tersebut membuktikan efektivitas TF-IDF pada teks kesehatan, meskipun belum membandingkan SVM dan LR.

5. Penelitian lain membandingkan XGBoost berbasis TF-IDF dengan IndoBERT pada 10.000 data tanya jawab kesehatan berbahasa Indonesia menggunakan validasi silang 5-fold. Hasil menunjukkan IndoBERT mencapai akurasi lebih tinggi, namun XGBoost menawarkan efisiensi komputasi yang jauh lebih baik dengan akurasi yang tetap kompetitif [11]. Penelitian ini relevan karena menggunakan data kesehatan berbahasa Indonesia, tetapi berfokus pada komparasi model klasik dan transformer.

6. Beberapa penelitian menunjukkan efektivitas SVM dan LR dengan ekstraksi fitur TF-IDF [7] [8], di antaranya perbandingan LR dan SVM pada 1.677 ulasan aplikasi PDAM berbasis TF-IDF dengan *hyperparameter tuning* pencarian acak menemukan bahwa kinerja kedua algoritma meningkat hingga akurasi 88% dan F1-*score* 83% [12]. Penelitian lain menerapkan Stacking Ensemble dengan SVM sebagai *base learner* dan LR sebagai meta *learner* pada 99.000 ulasan Ruangguru berbasis TF-IDF, dengan LR secara individual mengungguli algoritma lainnya [13].

7. Berdasarkan penelitian tersebut, dapat disimpulkan bahwa SVM dan Logistic Regression (LR) terbukti efektif dalam klasifikasi teks pada

Infotek : Jurnal Informatika dan Teknologi
Vol. 9 No. 2, Juli 2026
Hal.651-662

e-ISSN 2614-8773

DOI : 10.29408/jit.v9i2.35011 Link : https://dx.doi.org/10.29408/jit.v9i2.35011

berbagai jenis dataset. Selain itu, penggunaan TF-IDF sebagai representasi fitur telah banyak diterapkan pada klasifikasi teks kesehatan dan menunjukkan performa yang baik ketika dipadukan dengan algoritma *machine learning.* Namun, sebagian besar penelitian sebelumnya masih berfokus pada data rekam medis [9], data ulasan aplikasi [12][13], atau media sosial[8], sehingga belum merepresentasikan karakteristik teks konsultasi kesehatan daring yang bersifat naratif, tidak terstruktur, dan kontekstual.

8. Sementara itu, penelitian pada teks konsultasi medis dari *platform Online Health Consultation* (OHC) berbahasa Indonesia membuktikan bahwa TF-IDF lebih unggul dari Word2Vec dalam klasifikasi teks medis berbahasa Indonesia [14]. Hasil tersebut menunjukkan pendekatan berbasis TF-IDF memiliki potensi besar untuk diterapkan pada domain kesehatan reproduksi wanita yang masih belum banyak diteliti.

2.2. Landasan Teori

1. Telemedis

Telemedis adalah layanan kesehatan berbasis telekomunikasi untuk konsultasi pasien, melalui konferensi video atau aplikasi kesehatan [1].

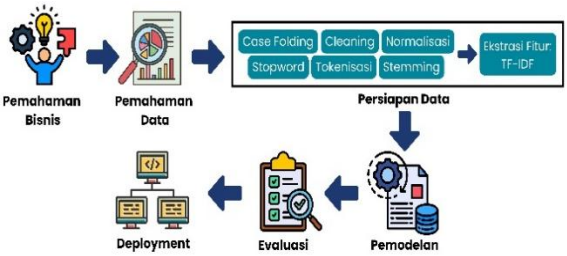
2. Data Mining

Data mining merupakan penggalian informasi dari kumpulan data besar menggunakan teknik statistik dan kecerdasan buatan sebagai dasar pengambilan keputusan [9]

3. Metode Penelitian

Cross Industry Standard Process for Data Mining (CRISP-DM)

Penelitian ini menerapkan metode *Cross Industry Standard Process for Data Mining* (CRISP-DM) seperti ditampilkan pada Gambar 1 [13], dengan tahapan sebagai berikut:



Gambar 1. Metode Penelitian

1. Pemahaman Bisnis

Penelitian ini mengklasifikasikan teks jawaban dokter pada konsultasi daring bertopik kehamilan dan menstruasi yang memiliki kosakata dan konteks serupa, tetapi memerlukan penanganan klinis yang berbeda. Untuk itu, penelitian membandingkan Support Vector Machine (SVM) dan Logistic Regression (LR) serta menganalisis istilah medis pembeda kedua topik.

2. Pemahaman Data

Tahap ini menggunakan dataset publik *"Doctor's Answer Text Dataset in Indonesian"* [15] yang terdiri atas 497.974 data hasil ekstraksi riwayat konsultasi daring Alodokter.com periode 8 Desember 2014–28 Februari 2021. Kolom *answer* digunakan sebagai fitur input, sedangkan *topic* sebagai label target.

Infotek : Jurnal Informatika dan Teknologi
Vol. 9 No. 2, Juli 2026
Hal.651-662

e-ISSN 2614-8773

DOI : 10.29408/jit.v9i2.35011 Link : https://dx.doi.org/10.29408/jit.v9i2.35011

Tabel 1. Contoh Dataset dengan Label

Answer	Label
Hallo, letak janin kep ini mungkin maksudnya kepala janin letaknya dibawah dan ini merupakan hal yang baik dan usia 25 minggu kehamilan biasanya tinggi fundus (bagian rahim paling puncak) ada 2 jari diatas pusar. Sekian, semoga bermanfaat yaa. Salam, dr.	0
Hai,Tidak ada hubungan antara keramas dan menstruasi . hal tersebut hanyalah mitos yang beredar di masyarakat. Anda tetap dapat keramas, mandi dan melakukan berbagai hal selama menstruasi.Anda dapat membaca artikel di bawah ini: Yang terjadi selama siklus menstruasi Semoga membantu dr.	1

Berdasarkan hasil filtrasi topik konsultasi pada dataset, topik kehamilan dan menstruasi memiliki jumlah konsultasi terbanyak, sehingga keduanya dipilih sebagai objek klasifikasi penelitian ini.

3. Persiapan Data

Tahap persiapan data mengolah data mentah sebelum pemodelan melalui proses:

a. Tahap Seleksi dan Pelabelan Data

Seleksi data dilakukan pada topik kehamilan dan menstruasi dengan pelabelan biner, yaitu 0 untuk kehamilan dan 1 untuk menstruasi. Dua data yang memiliki lebih dari satu label dihapus sehingga diperoleh 8.915 data kehamilan (50,07%) dan 8.890 data menstruasi (49,93%). Tabel 1 menunjukkan contoh dataset setelah pelabelan.

b. Tahapan Pra-Pemrosesan Data

Pra-pemrosesan teks jawaban dokter pada konsultasi daring dirancang sistematis, meliputi:

- Pembersihan *encoding* menggunakan ftfy untuk memperbaiki mojibake, normalisasi unicode NFKD, serta menghapus karakter non-ASCII.
- Penghapusan salam, sapaan, dan nama pasien di awal teks menggunakan regular expression (regex) bertingkat sebelum *case folding* karena ragam salam umumnya diawali huruf kapital.
- Seluruh teks diubah menjadi huruf kecil melalui *case folding*.
- Tahap *cleaning* teks untuk menghapus tag HTML, URL, angka, tanda baca, simbol, hingga kata tunggal satu huruf.
- Penghapusan *noise* untuk ucapan kalimat pembuka terima kasih, perkenalan dokter, kalimat penutup, serta baris *redirect*.
- Selanjutnya, penghapusan 5 data duplikat konflik dengan label berbeda dan 41 data duplikat dengan label yang sama.
- Normalisasi kata menggunakan kamus pemetaan (*mapping dictionary*) untuk frasa multi-kata, kata tunggal, singkatan informal, dan bahasa tidak baku. Proses dilakukan secara berurutan dan bersifat *case-insensitive*.
- Tokenisasi untuk memecah teks menjadi token individual dengan word_tokenize dari NLTK.
- Tahap ini menggabungkan *stopword* PySastrawi dengan *stopword custom* yang meliputi sisa *template platform*, sisa sapaan, kata medis tidak diskriminatif karena muncul hampir sama banyak pada kedua label, dan kata umum non-informatif.
- Proses *stemming* menggunakan PySastrawi dengan *exception list* berisi 16 istilah medis

Infotek : Jurnal Informatika dan Teknologi
Vol. 9 No. 2, Juli 2026
Hal.651-662

e-ISSN 2614-8773

DOI : 10.29408/jit.v9i2.35011 Link : https://dx.doi.org/10.29408/jit.v9i2.35011

sehingga kosakata berkurang 27.4% tanpa menghilangkan informasi medis yang bermakna. Setelah pra-pemrosesan, jumlah data berkurang dari 17.805 menjadi 17.703 yang siap digunakan. Perbandingan sampel data sebelum dan sesudah pra-pemrosesan ditampilkan pada Tabel 2.

Tabel 2. Sampel Data Sesudah Pra-pemrosesan

Sebelum Pra-pemrosesan	Setelah Pra-pemrosesan
Hai,Tidak ada hubungan antara keramas dan menstruasi . hal tersebut hanyalah mitos yang beredar di masyarakat. Anda tetap dapat keramas, mandi dan melakukan berbagai hal selama menstruasi.Anda dapat membaca artikel di bawah ini: Yang terjadi selama siklus menstruasi Semoga membantu dr.	['hubung', 'keramas', 'menstruasi', 'mitos', 'edar', 'masyarakat', 'keramas', 'mandi', 'menstruasi', 'baca', 'siklus', 'menstruasi']

4. Pemodelan

a. Model Latih dan Uji

Dataset dibagi menjadi data latih (*training set*) dan data uji (*testing set*) dengan empat variasi rasio, yaitu 90:10, 80:20, 70:30, dan 60:40 untuk menganalisis pengaruh proporsi data terhadap performa model klasifikasi. Hasil pembagian data ditampilkan pada Tabel 3.

Tabel 3. Hasil Pembagian Data

Rasio	Jumlah Data Train	Jumlah Data Test
90:10	15932	1771
80:20	14162	3541
70:30	12392	5311
60:40	10621	7082

b. Ekstraksi Fitur

Pada penelitian ini, ekstraksi fitur menggunakan metode *Term Frequency–Inverse Document Frequency* (TF-IDF) untuk mengubah teks hasil pra-pemrosesan menjadi representasi numerik, dengan bobot setiap kata dihitung melalui [16]:

$$TF - DF(i,j) = TF(i,j) \times IDF(i) \quad (1)$$

$$TF(i,j) = \frac{n(i,j)}{\sum_{k=1}^{q} n(k,j)} \quad (2)$$

$$IDF(i) = \log \frac{N}{1+df(i)} \quad (3)$$

Secara matematis, $n_{(i,j)}$ adalah jumlah kemunculan t_i dalam D_j, N adalah total jumlah dokumen, dan $df_{(i)}$ adalah jumlah dokumen yang mengandung t_i. Proses ekstraksi fitur menggunakan parameter max_features=15000 untuk merepresentasikan istilah medis yang relevan dari kedua topik, sekaligus menyaring *noise* untuk mengurangi beban komputasi tanpa menghilangkan makna penting yang dibutuhkan model, sehingga seluruh variasi rasio pembagian data menghasilkan 15.000 fitur yang seragam sebagai *input* bagi model klasifikasi.

c. Model Algoritma SVM dan LR

Penelitian ini menggunakan SVM dan Logistic Regression (LR). SVM diimplementasikan melalui LinearSVC yang menggunakan kernel linear untuk memperoleh *hyperplane* optimal dalam memisahkan dua kelas [9]. Metode ini efektif untuk data teks TF-IDF berdimensi tinggi dan bersifat *linearly separable* tanpa memerlukan karnel kompleks. Model ini menggunakan Scikit-learn dengan parameter C=1.0 untuk menjaga keseimbangan kompleksitas dan mencegah *overfitting* maupun *underfitting*. Selain itu,

Infotek : Jurnal Informatika dan Teknologi
Vol. 9 No. 2, Juli 2026
Hal.651-662

e-ISSN 2614-8773

DOI : 10.29408/jit.v9i2.35011 Link : https://dx.doi.org/10.29408/jit.v9i2.35011

digunakan max_iter=2000 untuk memastikan proses optimasi mencapai konvergensi, dengan *loss function squared hinge* dan regularisasi L2. Perhitungan LinearSVC sebagai berikut:

$$min\frac{1}{2}||w||^2 + C\sum_{i=1}^{n}\xi_i \quad (4)$$

dengan batasan

$$y_i(\omega \cdot x_i + b) \geq 1 - \xi_i,\ \xi_i \geq 0,\ \forall i \quad (5)$$

Dalam model ini, w adalah vektor bobot, b sebagai bias, C parameter regularisasi penyeimbang *maximum margin* dan kesalahan klasifikasi, dan ξi adalah variabel *slack* yang memungkinkan data tidak terpisahkan secara linear. Algoritma kedua yang digunakan adalah LR yang menggunakan fungsi sigmoid untuk menghitung probabilitas data terhadap kelas tertentu, menghasilkan nilai antara 0 dan 1 yang dikonversi menjadi prediksi kelas berdasarkan threshold [8]. Model ini dibangun dengan parameter C=1.0 sebagai nilai regularisasi standar untuk menjaga keseimbangan antara kemampuan mempelajari pola data dan mencegah *overfitting*. Selain itu, digunakan max_iter=1000 untuk memastikan *solver* lbfgs mencapai konvergensi stabil, class_weight= 'balanced' untuk mencegah bias kelas mayoritas melalui pembobotan proporsional, dan regularisasi L2 sebagai pengaturan default.

5. Evaluasi

Tahap evaluasi dilakukan untuk mengukur dan membandingkan kinerja kedua model menggunakan empat metrik, yaitu akurasi (*accuracy*), presisi (*precision*), *recall*, dan F1-*score*. Akurasi mengukur total prediksi benar, presisi menilai ketepatan kelas, *recall* menghitung kemampuan mendeteksi suatu kelas, dan F1-*score* merupakan rata-rata harmonik *precision* dan *recall* [14]. Berikut rumus evaluasi yang digunakan:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (6)$$

$$Precision = \frac{TP}{TP+FP} \quad (7)$$

$$Recall = \frac{TP}{TP+FN} \quad (8)$$

$$F1-score = 2 \times \frac{Precision \times Recall}{Precision+Recall} \quad (9)$$

Dengan TP menunjukkan data positif yang diprediksi benar, FP data negatif yang diprediksi positif, TN data negatif yang diprediksi benar, sedangkan FN data positif yang diprediksi negatif. Selain evaluasi metrik, uji *McNemar* digunakan sebagai metode statistik non-parametrik untuk mengukur signifikansi perbedaan performa kedua model pada data uji yang sama [17]. Pengujian dilakukan dengan menyusun matriks kontingensi 2×2 untuk memetakan koordinat prediksi kedua model. Nilai b menunjukkan jumlah data yang salah oleh SVM namun benar oleh LR, sedangkan nilai c menunjukkan kondisi sebaliknya. Statistik uji McNemar dihitung dengan rumus:

$$\chi^2 = \frac{(b-c)^2}{b+c} \quad (10)$$

Nilai Chi-Square (χ^2) dibandingkan dengan nilai kritis pada distribusi *chi-square* dengan 1 derajat

Infotek : Jurnal Informatika dan Teknologi
Vol. 9 No. 2, Juli 2026
Hal.651-662

e-ISSN 2614-8773

DOI : 10.29408/jit.v9i2.35011 Link : https://dx.doi.org/10.29408/jit.v9i2.35011

kebebasan signifikansi α = 0.05. Apabila χ^2 lebih besar dari nilai kritis, maka terdapat perbedaan performa yang signifikan antara kedua model.

Penelitian ini juga melakukan analisis *feature importance* untuk mengidentifikasi istilah yang paling berpengaruh dalam klasifikasi teks jawaban dokter menggunakan model dan rasio data terbaik. Signifikansi suatu fitur ditentukan oleh besarnya koefisien (β), yang dirumuskan sebagai berikut:

$$g(X) = \text{sigmoid}(\alpha + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n) \quad (11)$$

Dengan α sebagai intercept (bias), β_i dan x_i merupakan koefisien dan nilai TF-IDF kata ke-i.

6. Deployment

Tahap ini menginterpretasikan hasil evaluasi performa model dengan membandingkan SVM dan Logistic Regression (LR). Model terbaik kemudian dianalisis menggunakan metode *feature importance* untuk mengidentifikasi istilah medis yang paling berpengaruh dalam membedakan teks jawaban dokter pada konsultasi dengan topik kehamilan dan menstruasi.

4. Hasil dan Pembahasan

4.1. Hasil Penelitian

1. Visualisasi Distribusi Kata

Gambar 2 dan 3 menyajikan *word cloud* kelas Kehamilan (0) dan Menstruasi (1) yang menunjukkan perbedaan kata dominan, meskipun beberapa kata muncul pada kedua kelas dan berpotensi memengaruhi klasifikasi.

Gambar 2. Word Could Label Kehamilan

Gambar 3. Word Could Label Menstruasi

2. Hasil Ekstraksi Fitur TF-IDF

Pada tahap ini diperoleh 17.703 sampel dengan 15.000 fitur hasil ekstraksi TF-IDF.

3. Hasil Evaluasi Performa Model

Model Support Vector Machine (SVM)

Berdasarkan Tabel 4, SVM menunjukkan performa yang tinggi dan stabil pada seluruh rasio data, dengan akurasi terbaik pada rasio 80:20 sebesar 0.9305 yang menunjukkan bahwa komposisi data latih dan data uji pada rasio tersebut memberikan hasil pembelajaran yang paling efektif.

Tabel 4. Hasil Evaluasi Performa Support Vector Machine (SVM)

Rasio	Label	Precision	Recall	F1-Score	Akurasi
90:10	0	0.9269	0.9290	0.9279	0.9277
	1	0.9286	0.9265	0.9275	
80:20	0	0.9396	0.9205	0.9299	0.9305
	1	0.9218	0.9406	0.9311	
70:30	0	0.9354	0.9248	0.9301	0.9304
	1	0.9254	0.9359	0.9306	
60:40	0	0.9313	0.9250	0.9281	0.9283
	1	0.9253	0.9316	0.9284	

Infotek : Jurnal Informatika dan Teknologi
Vol. 9 No. 2, Juli 2026
Hal.651-662

e-ISSN 2614-8773

DOI : 10.29408/jit.v9i2.35011 Link : https://dx.doi.org/10.29408/jit.v9i2.35011

angka ditebalkan nilai tertinggi pada tiap kelas dan metrik. Nilai *precision* tertinggi kelas Kehamilan (0) diperoleh pada rasio 80:20 (0.9396), sedangkan kelas Menstruasi (1) pada rasio 90:10 (0.9286). Nilai *recall* tertinggi dicapai kelas Menstruasi (1) pada rasio 80:20 (0.9406) dan kelas Kehamilan (0) pada rasio 90:10 (0.9290). Sementara itu, F1-Score tertinggi kelas Kehamilan (0) diperoleh pada rasio 70:30 (0.9301) dan kelas Menstruasi (1) pada rasio 80:20 (0.9311). Variasi rasio terbaik pada setiap metrik menunjukkan bahwa performa tiap kelas dipengaruhi oleh keragaman kosakata pada kedua topik. Rasio 80:20 memberikan performa paling stabil dengan seluruh metrik di atas 0.92, sehingga SVM mampu mengklasifikasikan kedua topik secara akurat.

4. Model Logistic Regression

Pada Tabel 5, model LR menunjukkan performa yang konsisten pada seluruh rasio pengujian, dengan hasil terbaik pada rasio 80:20 yang menghasilkan akurasi 0.9359, *precision* tertinggi pada kedua kelas, serta F1-score sebesar 0.9358 (Kehamilan) dan 0.9360 (Menstruasi). Nilai *recall* tertinggi diperoleh kelas Kehamilan (0) pada rasio 90:10 (0.9369) dan kelas Menstruasi (1) pada rasio 80:20 (0.9383). Meskipun demikian, rasio 60:40 tetap memberikan F1-score yang seimbang (0.9318) pada kedua kelas, menunjukkan kestabilan performa LR meskipun menggunakan data latih yang lebih sedikit.

Tabel 5. Hasil Evaluasi Performa Logistic Regression (LR)

Rasio	Label	Precision	Recall	F1-Score	Akurasi
90:10	0	0.9275	0.9369	0.9321	0.9317
	1	0.9360	0.9265	0.9312	
80:20	0	0.9382	0.9334	0.9358	0.9359
	1	0.9336	0.9383	0.9360	
70:30	0	0.9349	0.9293	0.9321	0.9322
	1	0.9295	0.9351	0.9323	
60:40	0	0.9335	0.9301	0.9318	0.9318
	1	0.9301	0.9335	0.9318	

angka ditebalkan nilai tertinggi pada tiap kelas dan metrik Secara keseluruhan, seluruh rasio menghasilkan nilai evaluasi di atas 0.93 dengan selisih akurasi antar rasio yang sangat kecil, menunjukkan bahwa model LR memiliki performa yang stabil terhadap variasi pembagian data dalam klasifikasi teks jawaban dokter pada konsultasi kesehatan daring topik kehamilan dan menstruasi.

4.2. Pembahasan

1. Perbandingan Performa Tertinggi Kedua Model

Pada sebagian besar rasio, *Logistic Regression* (LR) menunjukkan performa sedikit lebih baik dibandingkan SVM dengan akurasi tertinggi 0.9359 dibandingkan 0.9305 pada rasio 80:20.

Berdasarkan Tabel 6, SVM unggul pada kelas Kehamilan (0) (0.9396), sedangkan LR unggul pada kelas Menstruasi (1) (0.9360), yang menunjukkan keunggulan masing-masing model pada kelas yang berbeda.

Tabel 6. Perbandingan Performa Precision

Infotek : Jurnal Informatika dan Teknologi
Vol. 9 No. 2, Juli 2026
Hal.651-662

e-ISSN 2614-8773

DOI : 10.29408/jit.v9i2.35011 Link : https://dx.doi.org/10.29408/jit.v9i2.35011

Label	Precision			
	Rasio	SVM	Rasio	LR
0	80:20	0.9396	80:20	0.9382
1	90:10	0.9286	90:10	0.9360

angka ditebalkan nilai tertinggi pada tiap kelas Pada Tabel 7, SVM unggul pada *recall* kelas Menstruasi (1) (0.9406), sedangkan LR unggul pada kelas Kehamilan (0) (0.9369), yang menunjukkan tidak ada model yang konsisten dominan pada kedua kelas untuk metrik *recall*.

Tabel 7. Perbandingan Performa Recall

Label	Recall			
	Rasio	SVM	Rasio	LR
0	90:10	0.9290	90:10	**0.9369**
1	80:20	**0.9406**	80:20	0.9383

angka ditebalkan nilai tertinggi pada tiap kelas

Tabel 8. Perbandingan Performa F1-Score

Label	F1-Score			
	Rasio	SVM	Rasio	LR
0	70:30	0.9301	80:20	**0.9358**
1	80:20	0.9311	80:20	**0.9360**

angka ditebalkan nilai tertinggi pada tiap kelas

Berdasarkan Tabel 8, LR menunjukkan performa terbaik dengan F1-*score* tertinggi pada kedua kelas pada rasio 80:20, yaitu 0.9358 untuk Kehamilan (0) dan 0.9360 untuk Menstruasi (1), sehingga lebih unggul dibandingkan SVM dalam klasifikasi teks jawaban dokter pada konsultasi kesehatan daring.

2. Pengujian Statistik McNemar

Hasil uji McNemar pada model klasifikasi SVM dan Logistic Regression (LR) dengan rasio terbaik 80:20 disajikan pada Tabel 9, menunjukkan bahwa 3.245 sampel yang diprediksi benar oleh kedua model, 55 sampel hanya diprediksi benar oleh LR, 39 sampel hanya diprediksi benar oleh SVM, dan 202 sampel yang diprediksi salah oleh keduanya.

Tabel 9. Matriks Kontingensi McNemar

	SVM Benar	SVM Salah
LR Benar	3245	55
LR Salah	39	202

Pada Tabel 10 menunjukkan nilai b + c = 94 (≥ 25), menggunakan Chi-Square dengan koreksi kontinuitas, menghasilkan nilai statistik sebesar 2.3936 dan p-value 0.1218. Dengan *p-value* (0.1218) > α (0.05), tidak ada perbedaan performa yang signifikan secara statistik antara LR dan SVM, sehingga pemilihan LR sebagai model terbaik didasarkan pada konsistensi performa yang lebih tinggi pada seluruh metrik evaluasi.

Tabel 10. Hasil Uji McNemar

Komponen	Nilai
b (LR Benar, SVM Salah)	55
c (LR Salah, SVM Benar)	39
b + c	94
Statistik McNemar	2.3936
p-value	0.1218
Tingkat Signifikansi (α)	0.05

3. Feature Importance

Analisis Feature Importance pada model LR terbaik (rasio 80:20) mengidentifikasi istilah medis pembeda topik kehamilan dan menstruasi.

Berdasarkan *feature importance* pada Tabel 11, model LR berhasil mengidentifikasi istilah medis yang paling berpengaruh dan representatif pada

Infotek : Jurnal Informatika dan Teknologi
Vol. 9 No. 2, Juli 2026
Hal.651-662

e-ISSN 2614-8773

DOI : 10.29408/jit.v9i2.35011 Link : https://dx.doi.org/10.29408/jit.v9i2.35011

setiap kategori. Hasil ini menunjukkan bahwa kombinasi TF-IDF dan LR tidak hanya menghasilkan performa klasifikasi yang baik, tetapi juga memberikan interpretasi yang jelas terhadap istilah yang membedakan teks jawaban dokter pada topik kehamilan dan menstruasi.

Tabel 11. Hasil Feature Importance dan Koefisien menggunakan Algoritma Logistic Regression

Label	Feature Importance dan Koefisien
0	hamil: -10.7327, janin: -5.5685, usia hamil: -3.2229, bayi: -3.1622, sperma: -3.1223, trimester: -2.9468, persalinan: -2.4743 wanita hamil: -2.1019, usia: -2.0226, buah: -1.8555, ketuban: -1.8103, perkembangan: -1.7864, hamil minggu: -1.6511, cepat hamil: -1.6122, plasenta: -1.5931, hamil wanita: -1.5751, minggu: -1.5654, hamil sebab: -1.5065, posisi: -1.4368, darah hamil: -1.4104
1	menstruasi: 11.4589, siklus: 4.2117, siklus menstruasi: 3.0281, darah menstruasi: 2.9697, menstruasi normal: 2.9111, menstruasi atur: 2.5942, hormonal: 2.5234, ganggu: 2.5194, stress: 2.4438, nifas: 2.3790, darah: 2.3010, atur: 2.2119, hormon: 2.1631, nyeri menstruasi: 2.1624, sebab: 2.1298, normal: 2.1042, olahraga: 2.0714, stres: 2.0377, ovarium: 1.9191, berat badan: 1.8952

4. Deployment

Tahap ini menghasilkan model klasifikasi otomatis untuk membedakan topik kehamilan dan menstruasi pada teks jawaban dokter. Berdasarkan hasil evaluasi, Logistic Regression (LR) berbasis TF-IDF memberikan performa terbaik sehingga direkomendasikan untuk klasifikasi konsultasi kesehatan daring. Model ini mendukung kategorisasi konsultasi yang lebih cepat dan efisien, sementara analisis *feature importance* menjadi dasar pengembangan rekomendasi informasi kesehatan dan chatbot medis berbahasa Indonesia yang lebih kontekstual

5. Kesimpulan

Penelitian ini membandingkan Support Vector Machine (SVM) dan Logistic Regression (LR) berbasis ekstraksi fitur TF-IDF untuk mengklasifikasikan teks jawaban dokter pada konsultasi kesehatan daring bertopik kehamilan dan menstruasi. Kedua model menunjukkan performa stabil pada seluruh rasio data, dengan rasio 80:20 sebagai konfigurasi terbaik. LR memberikan performa terbaik dengan akurasi 0.9359, lebih tinggi dibandingkan SVM (0.9305). Analisis *feature importance* mengidentifikasi istilah "hamil" dan "janin" sebagai fitur dominan pada kategori kehamilan, serta "menstruasi" dan "siklus" pada kategori menstruasi. Temuan ini berpotensi mendukung pengembangan sistem kategorisasi otomatis pada platform telemedis, seperti Smart-Triage dan Auto-Routing. Penelitian ini masih terbatas pada dua topik kesehatan reproduksi dari satu platform konsultasi serta menggunakan representasi TF-IDF yang belum sepenuhnya menangkap hubungan semantik antar kata.

6. Daftar Pustaka

[1] A. S. Sitanggang, R. G. Imanuel, N. I. Rapa, W. Shandie, and I. J. Halim, “Telemedicine : Revolusi Akses dan Efisiensi Pelayanan Kesehatan di Era

Infotek : Jurnal Informatika dan Teknologi
Vol. 9 No. 2, Juli 2026
Hal.651-662

e-ISSN 2614-8773

DOI : 10.29408/jit.v9i2.35011 Link : https://dx.doi.org/10.29408/jit.v9i2.35011

Digital," *Nusant. J. Multidiscip. Sci.*, vol. 2, no. 1, pp. 12–18, 2024.

[2] P. M. S. Rizqi, M. I. Hannari, S. Dewi, E. Alvionita, A. D. Shafira, and S. H. Purba, "Literatur Review : Transformasi Layanan Kesehatan Melalui Telemedicine : Peluang Dan Tantangan Di Era Digital," *J. Kesehat. Tambusai*, vol. 6, no. 1, pp. 699–707, 2025.

[3] D. P. Artiray, Misrawati, and W. P. Suci, "Gambaran Penggunaan Aplikasi Kesehatan Mobile dan Pengetahuan Ibu Hamil tentang Kehamilan," *J. Ners Univ. Pahlawan*, vol. 9, no. 1, pp. 1216–1224, 2025.

[4] T. N. Putri, "Peran Aplikasi Mobile Health Dalam Pemantauan Siklus Menstruasi," *J. Innov. Creat.*, vol. 5, no. 3, pp. 34654–34663, 2025.

[5] E. Winengsih, T. Lubis, K. Khopiati, and N. N. Rahmah, "Deteksi Kesehatan Mental dengan Gangguan Menstruasi Remaja Gen Z," *J. Kesehat. Perintis*, vol. 11, no. 2, pp. 159–167, 2025.

[6] N. E. C. Dewi and V. N. Hafifah, "Edukasi eMAMA dalam Meningkatkan Pengetahuan Ibu Hamil terhadap faktor Risiko Kematian Maternal," *Indones. Heal. Sci. J.*, vol. 5, no. 2, pp. 122–128, 2025.

[7] N. W. A. Prasetya, L. P. Wanti, and R. Purwanto, "Studi Perbandingan Kinerja Support Vector Machine Pada Klasifikasi Diabetes Mellitus Menggunakan Fitur Regular Expression dan Non- Regular Expression," *Infotekmesin*, vol. 17, no. 1, pp. 190–198, 2026.

[8] A. Firizkiansah, A. Muhammad, and I. R. Maulana, "Optimasi Klasifikasi Data Teks Menggunakan Algoritma Logistic Regression dengan TF-IDF dan SMOTE," *JIKOMTI J. Ilm. Ilmu Komput. dan Teknol. Inf.*, vol. 2, no. 1, pp. 29–38, 2025.

[9] G. Ben Abdennour, K. Gasmi, and R. Ejbali, "An Optimal Model for Medical Text Classification Based on Adaptive Genetic Algorithm," *Data Sci. Eng.*, vol. 9, no. 4, pp. 378–392, 2024.

[10] M. J. Faisti, R. H. Kusumodestoni, and G. W. N. Wibowo, "Mental Health Classification Using Naïve Bayes and Random Forest Algorithms," *J. Appl. Informatics Comput.*, vol. 9, no. 4, pp. 1740–1750, 2025.

[11] A. G. S. Moeslim, E. Firmansyah, and B. Sutara, "Perbandingan Kinerja XGBoost dan IndoBERT untuk Klasifikasi Teks Kesehatan Bahasa Indonesia," *Data Sci. Indones.*, vol. 5, no. 2, pp. 62–72, 2025.

[12] S. A. Ramadani, H. Hamrul, and N. Arifin, "Analisis Pengaruh Random Search Pada Logistic Regression dan Support Vector Machine Dalam Klasifikasi Sentimen Aplikasi PDAM Info," *J. SKANIKA Sist. Komput. dan Tek. Inform.*, vol. 9, no. 1, pp. 61–75, 2026.

[13] D. Sofia and P. Sekarpuji, "Penerapan Metode Stacking Ensemble Untuk Analisis Sentimen Pada Ulasan Aplikasi RuangGuru," *Idealis Indones. J. Inf. Syst.*, vol. 8, no. 2, pp. 248–257, 2025.

[14] S. Juanita, D. Purwitasari, I. K. E. Purnama, A. F. Abdillah, and M. H. Purnomo, "Multi-Label Classification for Doctor's Behavioral Pattern Matching During Online Medical Interview using Machine Learning," *Int. J. Electr. Eng. Informatics*, vol. 15, no. 3, pp. 435–452, 2023.

[15] S. Juanita, D. Purwitasari, I. K. E. Purnama, and M. Purnomo, "Doctor's Answer Text Dataset in Indonesian Contains Information on Medical Interview Patterns," 2023..

[16] L. Zhang, "Features Extraction Based on Naive Bayes Algorithm and TF-IDF for News Classification," *PLoS One*, vol. 20, no. 7, pp. 1–17, 2025.

[17] O. Rainio, J. Teuho, and R. Klén, "Evaluation Metrics and Statistical Tests for Machine Learning," *Sci. Rep.*, vol. 14, no. 6068, pp. 1–14, 2024