

**PENGEMBANGAN *CHATBOT* KLASIFIKASI MITOS ATAU
FAKTA KESEHATAN UMUM BERBAHASA INDONESIA**

SKRIPSI

Diajukan Guna Memenuhi Salah Satu Syarat
Memperoleh Gelar Sarjana Komputer (S.Kom.)
Pada Program Studi Teknik Informatika



Oleh:

Rizky Ananda Putra
NPM: 2213020119

**FAKULTAS TEKNIK DAN ILMU KOMPUTER
UNIVERSITAS NUSANTARA PGRI KEDIRI**

2026

Skripsi oleh:

Rizky Ananda Putra
NPM: 2213020119

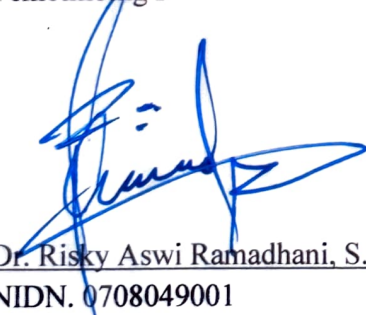
Judul:

**PENGEMBANGAN *CHATBOT* KLASIFIKASI MITOS ATAU
FAKTA KESEHATAN UMUM BERBAHASA INDONESIA**

Telah Disetujui untuk Diajukan Kepada Panitia Ujian/Sidang Skripsi
Program Studi Teknik Informatika Fakultas Teknik dan Ilmu Komputer
Universitas Nusantara PGRI Kediri

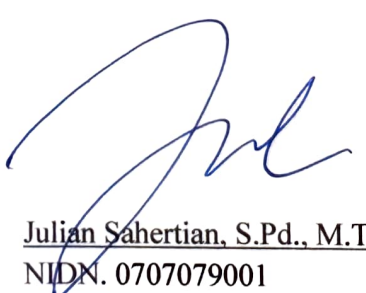
Tanggal: 19 Juni 2026

Pembimbing I



Dr. Risky Aswi Ramadhani, S.Kom., M.Kom.
NIDN. 0708049001

Pembimbing II



Julian Sahertian, S.Pd., M.T.
NIDN. 0707079001

Skripsi oleh:

Rizky Ananda Putra
NPM: 2213020119

Judul:

**PENGEMBANGAN *CHATBOT* KLASIFIKASI MITOS ATAU
FAKTA KESEHATAN UMUM BERBAHASA INDONESIA**

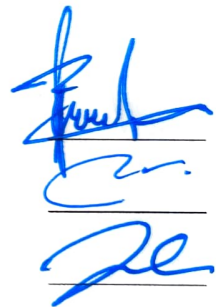
Telah dipertahankan di depan Panitia Ujian/Sidang Skripsi
Program Studi Teknik Informatika Fakultas Teknik dan Ilmu Komputer
Universitas Nusantara PGRI Kediri

Pada tanggal: 21 Juli 2026

Dan Dinyatakan telah Memenuhi Syarat

Panitia Penguji :

1. Ketua : Dr. Risky Aswi Ramadhani, S.Kom., M.Kom.
2. Penguji I : Ratih Kumalasari Niswatin, S.ST., M.Kom.
3. Penguji II : Julian Sahertian, S.Pd., M.T.



Mengetahui,
Dekan ETIK

Dr. Sahstiono, M.Si.
NIDN. 0007076801

HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini saya,

Nama : Rizky Ananda Putra
Jenis Kelamin : Laki-laki
Tempat/Tgl Lahir : Purworejo, 10 April 2004
NPM : 2213020119
Fakultas/Prodi : Teknik dan Ilmu Komputer/ Teknik Informatika

Menyatakan dengan sebenarnya, bahwa dalam Skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu perguruan tinggi, dan sepanjang pengetahuan saya tidak dapat karya tulis atau pendapat yang pernah diterbitkan oleh orang lain, kecuali yang secara sengaja dan tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Kediri, 19 Juni 2026
Yang Menyatakan



Rizky Ananda Putra
NPM 2213020119

HALAMAN PERSEMBAHAN

Dengan rasa syukur dan haru yang mendalam, skripsi ini penulis persembahkan sebagai wujud terima kasih kepada:

1. Kedua Orang Tua saya, terima kasih untuk setiap doa yang tak pernah putus, pengorbanan, dan kesabaran luar biasa yang menjadi kekuatan terbesar penulis untuk bangkit saat merasa lelah dalam menyelesaikan skripsi ini.
2. Kakak-kakak saya, atas segala bentuk bantuan, semangat, dan dukungan nyata yang selalu ada sehingga proses penyusunan tugas akhir ini dapat terselesaikan dengan baik.
3. Bapak Dr. Risky Aswi Ramadhani, S.Kom., M.Kom., dan Bapak Julian Sahertian, S.Pd., M.T., selaku Dosen Pembimbing Skripsi atas keikhlasan dalam meluangkan waktu, tenaga, serta kesabaran yang luar biasa dalam membimbing dan mengarahkan penulis ketika menemui kebuntuan.
4. Seluruh Dosen Program Studi Teknik Informatika, Universitas Nusantara PGRI Kediri, atas ilmu, wawasan, dan kesabaran dalam mendidik yang menjadi fondasi penting bagi penulis hingga sampai di tahap ini.
5. Teman-teman seperjuangan di kampus, terima kasih telah menjadi tempat berbagi beban, berkeluh kesah, dan saling menguatkan untuk menolak menyerah hingga titik akhir ini.
6. Almamater Tercinta Universitas Nusantara PGRI Kediri, yang telah menjadi tempat saya tumbuh, belajar, dan mengembangkan diri selama masa perkuliahan.

Serta semua pihak yang tidak dapat penulis sebutkan satu per satu, yang telah memberikan dukungan berarti selama proses penyusunan. Semoga dedikasi ini dapat menjadi bentuk penghormatan yang layak atas segala doa dan kebaikan yang telah diberikan.

HALAMAN MOTTO

الْحَمْدُ لِلَّهِ رَبِّ الْعَالَمِينَ

“Segala puji bagi Allah, Tuhan seluruh alam.”

— QS. Al-Fatihah [1]: 2

إِنَّا لِلَّهِ وَإِنَّا إِلَيْهِ رَاجِعُونَ

“Sesungguhnya kami milik Allah, dan sesungguhnya hanya kepada-Nya lah kami kembali.”

— Q.S. Al-Baqarah [2]: 156 (dikutip oleh Sitara dalam anime *Jaadugar: A Witch in Mongolia*)

رَبَّنَا مَا خَلَقْتَ هَذَا بَاطِلًا سُبْحَانَكَ قَوْلًا عَذَابَ النَّارِ

“Ya Tuhan kami, tidaklah Engkau menciptakan semua ini sia-sia. Maha Suci Engkau, lindungilah kami dari azab neraka”

— QS. Ali ‘Imran [3]: 191

إِنَّ صَلَاتِي وَنُسُكِي وَمَحْيَايَ وَمَمَاتِي لِلَّهِ رَبِّ الْعَالَمِينَ

“Sesungguhnya salatku, ibadahku, hidupku, dan matiku hanyalah untuk Allah, Tuhan seluruh alam.”

— QS. Al-An’am [6]: 162

وَأَنْتُمْ سَامِدُونَ . فَاسْجُدُوا لِلَّهِ وَاعْبُدُوا

“Sedangkan kamu lengah (dari-Nya). Maka bersujudlah kepada Allah dan sembahlah (Dia).”

— QS. An-Najm [53]: 61-62

RINGKASAN

Rizky Ananda Putra Pengembangan *Chatbot* Klasifikasi Mitos Atau Fakta Kesehatan Umum Berbahasa Indonesia, Skripsi, Program Studi Teknik Informatika, Fakultas Teknik dan Ilmu Komputer, Universitas Nusantara PGRI Kediri, 2026

Kata Kunci: Klasifikasi Teks, Mitos Kesehatan, *Support Vector Machine* (SVM), *Term Frequency-Inverse Document Frequency* (TF-IDF), *Chatbot Telegram*.

Tingginya volume pencarian informasi kesehatan di Indonesia yang diiringi oleh mudahnya peredaran hoaks memicu meluasnya mitos kesehatan di tengah masyarakat. Penelitian ini bertujuan untuk membangun *dataset* mitos dan fakta kesehatan umum berbahasa Indonesia yang divalidasi oleh ahli, serta mengimplementasikannya ke dalam sistem klasifikasi berbasis *chatbot* pada platform *Telegram*. Menggunakan desain Penelitian Pengembangan dengan model *Waterfall*, sebanyak 200 data mentah dikumpulkan secara kurasi manual dari berbagai situs web kesehatan terpercaya dan telah divalidasi kebenarannya oleh dokter umum. Untuk memperkaya variasi fitur teks, *dataset* diekspansi menjadi 400 entri melalui teknik augmentasi *back-translation*, kemudian diproses melalui tahapan normalisasi, tokenisasi, *stopwords removal*, *stemming* Sastrawi, dan pembobotan *Term Frequency-Inverse Document Frequency*. Klasifikasi dilakukan menggunakan algoritma *Support Vector Machine* yang dievaluasi melalui *Stratified K-Fold Cross Validation*. Hasil pengujian menunjukkan konfigurasi terbaik diperoleh pada kombinasi parameter *Grid Search* menggunakan fitur *Unigram* (*max_features* 800), kernel RBF, dan $C=0.5$, yang menghasilkan performa dengan akurasi 0.9500, serta *F1-score* sebesar 0.9000 untuk kelas Mitos dan 0.9667 untuk kelas Fakta. Pengujian generalisasi pada 20 data baru tervalidasi pakar mencatatkan akurasi 0.5500, yang menunjukkan keterbatasan fitur *Unigram* dalam menangkap semantik kata negasi dan kosa kata baru (*out-of-vocabulary*). Secara fungsional dan non-fungsional, *chatbot* terbukti valid secara *Black Box*, memiliki waktu respons di bawah 3 detik, serta mampu melakukan penolakan (*error handling*) terhadap masukan berformat non-teks, menjadikannya layak digunakan sebagai purwarupa media edukasi literasi kesehatan masyarakat.

PRAKATA

Puji Syukur dipanjatkan kepada Tuhan Yang Maha Esa, karena atas rida dan karuniaNya peneliti dapat menyelesaikan penyusunan laporan penelitian ini. Penulisan ini juga tak lepas dari dukungan pihak yang selalu membantu dalam penulisan penelitian ini. Oleh karenanya peneliti ingin mengucapkan terima kasih kepada

1. Dr. Zainal Affandi, M.Pd., selaku Rektor Universitas Nusantara PGRI Kediri, yang selalu memberikan dorongan motivasi kepada mahasiswa.
2. Dr. Sulistiono, M.Si., selaku Dekan Fakultas Teknik dan Ilmu Komputer, yang selalu memberikan dukungan moral kepada mahasiswa.
3. Risa Helilintar, S.Kom., M.Kom., selaku Ketua Program Studi Teknik Informatika, yang selalu memberikan arahan kepada mahasiswa.
4. Dr. Risky Aswi Ramadhani, S.Kom., M.Kom., dan Julian Sahertian, S.Pd., M.T., selaku Dosen Pembimbing Skripsi yang telah mengarahkan selama mengerjakan skripsi.
5. Kedua Orang Tua saya dan Keluarga atas doa dan dukungannya.
6. Ucapan terima kasih juga disampaikan kepada pihak-pihak lain yang tidak dapat disebutkan satu persatu, yang telah banyak membantu menyelesaikan penulisan penelitian ini.

Disadari penelitian ini masih banyak kekurangan, maka diharapkan kritik dan saran dari berbagai pihak. Semoga penelitian ini bermanfaat bagi semua pihak.

Kediri, 19 Juni 2026

Rizky Ananda Putra
NPM 2213020119

DAFTAR ISI

HALAMAN SAMPUL.....	i
LEMBAR PERSETUJUAN PEMBIMBING SKRIPSI.....	ii
LEMBAR PENGESAHAN	iii
LEMBAR PERNYATAAN KEASLIAN TULISAN	iv
HALAMAN PERSEMBAHAN	v
HALAMAN MOTTO	vi
HALAMAN RINGKASAN	vii
HALAMAN PRAKATA.....	viii
DAFTAR ISI.....	ix
DAFTAR TABEL	xi
DAFTAR GAMBAR.....	xiii
DAFTAR LAMPIRAN	xv
BAB I PENDAHULUAN.....	1
A. Latar Belakang	1
B. Identifikasi Masalah	2
C. Rumusan Masalah	3
D. Batasan Masalah.....	3
E. Tujuan Penelitian	4
F. Manfaat Penelitian	4
BAB II LANDASAN TEORI	6
A. Teori dan Penelitian Terdahulu.....	6
1. Landasan Teori.....	6
2. Kajian Pustaka.....	14
B. Kerangka Berpikir	22
BAB III METODOLOGI PENELITIAN	23
A. Desain Penelitian.....	23
B. Instrumen Penelitian.....	26
C. Tempat dan Jadwal Penelitian.....	29

D. Objek dan Subjek Penelitian	30
E. Prosedur Penelitian.....	32
F. Teknik Analisis Data.....	36
BAB IV HASIL DAN PEMBAHASAN	50
A. Hasil Penelitian	50
B. Pembahasan.....	98
BAB V PENUTUP	101
A. Kesimpulan	101
B. Saran.....	102
DAFTAR PUSTAKA	103
LAMPIRAN.....	110

DAFTAR TABEL

Tabel 2.1 <i>Confusion Matrix</i>	12
Tabel 2.2 Penelitian Terdahulu	14
Tabel 3.1 Spesifikasi Perangkat Keras.....	26
Tabel 3.2 Spesifikasi Perangkat Lunak.....	26
Tabel 3.3 Linimasa Jadwal Penelitian.....	30
Tabel 3.4 Contoh Data	45
Tabel 3.5 Perhitungan Bobot Kata pada Dokumen Uji	47
Tabel 4.1 Data Simulasi	55
Tabel 4.2 Simulasi Proses <i>Case Folding</i> dan <i>Cleaning Symbol</i>	56
Tabel 4.3 Simulasi Proses Tokenisasi	56
Tabel 4.4 Simulasi Proses <i>Stopwords Removal</i>	57
Tabel 4.5 Simulasi Proses <i>Stemming</i>	57
Tabel 4.6 Perhitungan Bobot TF-IDF pada Data <i>Input</i>	58
Tabel 4.7 Sampel <i>Dataset</i> Pernyataan Kesehatan yang Telah Divalidasi	60
Tabel 4.8 Skenario Pengujian Fungsional.....	62
Tabel 4.9 Skenario Pengujian Non-Fungsional	63
Tabel 4.10 Pengujian Rasio Data tanpa Augmentasi	65
Tabel 4.11 Sampel Data Pengujian Rasio Data tanpa Augmentasi	68
Tabel 4.12 Pengujian Rasio Data dengan Augmentasi	69

Tabel 4.13 Sampel Data Pengujian Rasio Data dengan Augmentasi.....	71
Tabel 4.14 Pengujian TF-IDF tanpa Augmentasi	73
Tabel 4.15 Sampel Data Pengujian TF-IDF tanpa Augmentasi.....	75
Tabel 4.16 Pengujian TF-IDF dengan Augmentasi	76
Tabel 4.17 Sampel Data Pengujian TF-IDF dengan Augmentasi.....	79
Tabel 4.18 Pengujian SVM tanpa Augmentasi	80
Tabel 4.19 Sampel Data Pengujian SVM tanpa Augmentasi	82
Tabel 4.20 Pengujian SVM dengan Augmentasi	83
Tabel 4.21 Sampel Data Pengujian SVM dengan Augmentasi.....	86
Tabel 4.22 Pengujian <i>Grid Search</i> tanpa Augmentasi	88
Tabel 4.23 Sampel Data Pengujian <i>Grid Search</i> tanpa Augmentasi	89
Tabel 4.24 Pengujian <i>Grid Search</i> dengan Augmentasi	91
Tabel 4.25 Sampel Data Pengujian <i>Grid Search</i> dengan Augmentasi.....	92
Tabel 4.26 Hasil Klasifikasi pada 20 Data Baru	95
Tabel 4.27 Hasil Evaluasi Metrik pada Pengujian Data Baru.....	97

DAFTAR GAMBAR

Gambar 2.1 Kerangka Berpikir	22
Gambar 3.1 Alur Pengumpulan Data Kesehatan	25
Gambar 3.2 Model Pengembangan <i>Waterfall</i>	32
Gambar 3.3 <i>Use Case Diagram</i>	36
Gambar 3.4 <i>Activity Diagram</i>	37
Gambar 3.5 <i>Sequence Diagram</i>	39
Gambar 3.6 <i>Class Diagram</i>	40
Gambar 3.7 Desain <i>Wireframe Chatbot</i>	43
Gambar 3.8 <i>Flowchart</i> Pemrosesan Data.....	44
Gambar 4.1 Tampilan Pesan Pembuka dan Panduan Sistem	51
Gambar 4.2 Tampilan Dialog Klasifikasi Teks.....	52
Gambar 4.3 Tampilan Penanganan Kesalahan Terhadap Berbagai Format <i>Input</i> Non-Teks	53
Gambar 4.4 Grafik Pengujian Rasio Data tanpa Augmentasi.....	66
Gambar 4.5 Grafik Pengujian Rasio Data dengan Augmentasi	70
Gambar 4.6 Grafik Pengujian TF-IDF tanpa Augmentasi	73
Gambar 4.7 Grafik Pengujian TF-IDF dengan Augmentasi	77
Gambar 4.8 Grafik Pengujian SVM tanpa Augmentasi.....	80
Gambar 4.9 Grafik Pengujian SVM dengan Augmentasi.....	84

Gambar 4.10 <i>Confusion Matrix</i> Hasil Pengujian <i>Grid Search</i> tanpa Augmentasi	
Data	90
Gambar 4.11 <i>Confusion Matrix</i> Hasil Pengujian <i>Grid Search</i> dengan Augmentasi	
Data	93
Gambar 4.12 <i>Confusion Matrix</i> Pengujian Data Baru	96
Gambar 4.13 Contoh <i>Input</i> yang Baik dan <i>Input</i> yang Buruk	100

DAFTAR LAMPIRAN

Lampiran 1. Lembar Bimbingan Skripsi.....	110
Lampiran 2. Surat Penelitian Skripsi	112
Lampiran 3. Surat Keterangan Validator	113
Lampiran 4. Lembar Angket Pengujian Fungsional	114
Lampiran 5. Lembar Angket Pengujian Non-Fungsional.....	116
Lampiran 6. Lembar Angket Pengujian Data Baru.....	118
Lampiran 7. Sampel <i>Dataset</i> yang Tervalidasi	120

BAB I

PENDAHULUAN

A. Latar Belakang

Perkembangan internet saat ini telah mengubah cara masyarakat umum dalam mengakses informasi (Yu dkk., 2024). Hal ini menjadikan media *online* memiliki persebaran informasi yang lebih cepat dibandingkan media konvensional (Pratama dkk., 2024). Diantara informasi tersebut, kesehatan menempati posisi pertama sebagai informasi yang paling dicari masyarakat Indonesia, di mana menurut data Asosiasi Penyelenggara Jasa Internet Indonesia (APJII) 36,96% responden mengaku mencari informasi tentang kesehatan (GoodStats, 2023).

Tingginya volume pencarian informasi kesehatan ini sayangnya dihadapkan pada sebuah realita, di mana tidak semua konten yang tersebar di internet memiliki kevalidan dari ahli tentang benar tidaknya informasi tersebut (Sitohang & Lestari, 2025). Dilansir dari Kementerian Komunikasi dan Digital Republik Indonesia, terdapat 1923 konten hoaks yang tersebar sepanjang tahun 2024. Konten tentang kesehatan bahkan menempati posisi ke-empat, dengan 163 konten hoaks (Kementerian Komunikasi dan Digital, 2025). Jika dibiarkan berkelanjutan, hal ini akan menimbulkan kebingungan di tengah masyarakat, membangun persepsi yang salah terhadap kesehatan, dan hal ini dapat menjadi permasalahan serius (Kožuh & Čakš, 2023).

Permasalahan serius ini seringkali berwujud sebagai mitos kesehatan, yaitu narasi atau teks keliru yang telah lama beredar dan menyebar luas di kalangan masyarakat (Woodhouse, 2023). Untuk menjawab permasalahan serius ini diperlukan sebuah sistem klasifikasi teks yang mampu membedakan mitos atau fakta (Schlicht dkk., 2024). Pada klasifikasi di penelitian ini difokuskan menggunakan *Support Vector Machine* (SVM). Metode SVM sendiri memiliki keunggulan dalam menangani data teks dan menemukan bidang pemisah (*hyperplane*) yang optimal antar kelas (Lubis & Putri, 2025).

Kefektifan metode SVM ini dapat dilihat dari riset yang dilakukan oleh M. Sudakhar dan K.P. Kaliyamurthie pada tahun 2024 dalam mendeteksi berita palsu dari media sosial menggunakan algoritma SVM yang mana mendapat akurasi tinggi sebesar 98% (Sudhakar & Kaliyamurthie, 2024). Hal ini didukung pula oleh penelitian oleh Karim dkk. (2024) dimana memanfaatkan algoritma SVM dan teknik vektorisasi teks yang lebih canggih sebagai alternatif dari model BERT. Diperoleh metode SVM dengan pendekatan TF-IDF mampu bersaing dengan mendapat akurasi sebesar 99,52%, sedangkan dengan pendekatan BoW mendapat akurasi sebesar 99,81%. Lebih lanjut, pada studi oleh Sikosana dkk. (2024), dimana SVM dengan pendekatan *lemmatization* mendapat performa akurasi sebesar 94,14% dalam klasifikasi misinformasi di bidang kesehatan menjadikannya sebuah pilihan metode yang solid dan relevan.

Berdasarkan uraian di atas, penelitian ini bertujuan untuk merancang dan membangun sebuah sistem klasifikasi mitos atau fakta di bidang kesehatan umum berbahasa Indonesia yang diimplementasikan dalam bentuk *chatbot* pada platform *Telegram*. Aplikasi *Telegram* dipilih karena karakteristiknya yang ramah pengguna (*user-friendly*) serta kemampuannya dalam meningkatkan efektivitas berbagi informasi, sehingga platform ini tetap mudah diakses oleh masyarakat luas (I. A. Mohammed dkk., 2024). Untuk mencapai tujuan tersebut, penelitian ini akan dimulai dari pengumpulan data teks yang disimpan ke dalam file *Excel* untuk anotasi (Wei & Zhang, 2024). Dilanjutkan dengan validasi data oleh ahli, *preprocessing* data, serta pemodelan menggunakan metode SVM dengan parameter dan pendekatan yang disesuaikan. Kinerja model klasifikasi nantinya akan dievaluasi menggunakan *Stratified Cross-Fold Validation* (Hasanati dkk., 2023) dan *Confusion Matrix* yang mencakup *accuracy*, *precision*, *recall*, dan *F1-Score* (Aziz dkk., 2024).

B. Identifikasi Masalah

Berdasarkan latar belakang yang telah diuraikan sebelumnya, maka identifikasi masalah yang didapatkan sebagai berikut:

1. Belum tersedianya *dataset* untuk klasifikasi mitos atau fakta di bidang kesehatan umum berbahasa Indonesia yang divalidasi oleh ahli.
2. Belum adanya sistem berbasis SVM yang dapat mengklasifikasikan pernyataan kesehatan umum berbahasa Indonesia sebagai mitos atau fakta yang terintegrasi dengan platform *Telegram*.

C. Rumusan Masalah

Berdasarkan identifikasi masalah di atas, maka rumusan masalah dari penelitian ini adalah:

1. Bagaimana membangun sebuah *dataset* mitos atau fakta di bidang kesehatan umum berbahasa Indonesia yang divalidasi oleh ahli?
2. Bagaimana mengembangkan *Chatbot Telegram* berbasis SVM yang mampu mengklasifikasikan pernyataan kesehatan umum berbahasa Indonesia menjadi mitos atau fakta berdasarkan *dataset* yang telah dibangun?

D. Batasan Masalah

Agar penelitian ini tetap terfokus dan tidak meluas, maka terdapat beberapa batasan masalah sebagai berikut:

1. Penelitian ini berfokus pada klasifikasi teks biner dalam bidang kesehatan umum, dengan dua kategori *output*, yaitu hanya mitos dan fakta.
2. Seluruh teks klaim kesehatan yang digunakan berbahasa Indonesia, berupa kalimat atau pernyataan yang merepresentasikan informasi kesehatan.
3. Data dikumpulkan dari situs terpercaya seperti Alodokter, Halodoc, situs web rumah sakit, dan sebagainya yang berbasis kesehatan di Indonesia dengan rentang tahun publikasi 2020–2025.
4. Label mitos atau fakta pada *dataset* divalidasi oleh satu praktisi medis (dokter) untuk memastikan akurasi *ground truth*.
5. Metode yang digunakan untuk klasifikasi hanyalah SVM dengan ekstraksi fitur TF-IDF, dan tidak membandingkan dengan metode lain.
6. Prototipe sistem dikembangkan dalam bentuk *Chatbot* pada platform *Telegram*.

7. Sistem hanya menerima *input* berupa teks melalui percakapan pribadi (*private chat*). *Input* berupa gambar, suara, stiker, atau penggunaan dalam grup tidak dicakup dalam sistem ini.
8. *Chatbot* hanya menampilkan hasil klasifikasi (Mitos/Fakta) berdasarkan *input* teks pengguna, tanpa memberikan penjelasan medis, saran pengobatan, ataupun tanggapan konsultasi.
9. Evaluasi terbatas yang hanya menggunakan *Confusion Matrix* dengan metrik *accuracy*, *precision*, *recall*, dan *F1-score*.

E. Tujuan Penelitian

Berdasarkan rumusan masalah yang telah disebutkan di atas, maka tujuan dari penelitian ini adalah:

1. Membangun *dataset* mitos atau fakta di bidang kesehatan umum berbahasa Indonesia yang divalidasi oleh ahli.
2. Mengembangkan *Chatbot Telegram* berbasis SVM untuk mengklasifikasikan pernyataan kesehatan umum berbahasa Indonesia menjadi mitos atau fakta dengan memanfaatkan *dataset* yang telah dibangun.

F. Manfaat Penelitian

1. Manfaat Teoritis:
 - a. Menyediakan sebuah *dataset* baru untuk klasifikasi mitos atau fakta di bidang kesehatan umum berbahasa Indonesia yang divalidasi oleh ahli, yang dapat digunakan sebagai pengujian dan pengembangan model klasifikasi teks kedepannya.
 - b. Memberikan kontribusi pada pengembangan literatur ilmiah terkait penerapan algoritma SVM dalam klasifikasi teks kesehatan berbahasa Indonesia, khususnya pada konteks pemisahan antara pernyataan mitos dan fakta dengan data yang divalidasi oleh ahli.

2. Manfaat Praktis:

- a. *Dataset* yang dihasilkan dapat dimanfaatkan secara praktis oleh pengembang, platform kesehatan, atau lainnya sebagai data latih untuk membangun atau meningkatkan akurasi alat deteksi mitos kesehatan.
- b. Dapat menjadi referensi awal bagi institusi kesehatan ataupun media dalam mengembangkan sistem otomatis untuk deteksi mitos kesehatan, terutama pada platform populer seperti *Telegram*.

DAFTAR PUSTAKA

- Abonizio, H. Q., Barbon, A. P. A. da C., Rodrigues, R., Santos, M., Martínez-Vizcaino, V., Mesas, A. E., & Barbon Junior, S. (2023). How people interact with a chatbot against disinformation and fake news in COVID-19 in Brazil: The CoronaAI case. *International Journal of Medical Informatics*, 177, 105134. <https://doi.org/https://doi.org/10.1016/j.ijmedinf.2023.105134>
- Adamopoulou, E., & Moussiades, L. (2020). Chatbots: History, technology, and applications. *Machine Learning with Applications*, 2, 100006. <https://doi.org/https://doi.org/10.1016/j.mlwa.2020.100006>
- Alshammari, M., Doody, O., & Richardson, I. (2018). Barriers to the Access and use of Health Information by Individuals with Intellectual and Developmental Disability IDD: A Review of the Literature. *2018 IEEE International Conference on Healthcare Informatics (ICHI)*, 294–298. <https://api.semanticscholar.org/CorpusID:50784379>
- Arifin, N., Enri, U., & Sulistiyowati, N. (2021). Penerapan Algoritma Support Vector Machine (SVM) dengan TF-IDF N-Gram untuk Text Classification. *STRING (Satuan Tulisan Riset dan Inovasi Teknologi)*, 6(2), 129–136.
- Auliana, S., Kalsum, U., & Mudofar, M. (2025). Detecting Hoax News with Web-Based SVM Classifier: A Machine Learning Approach to Combat Misinformation Using Support Vector Machine (SVM) Algorithm. *JINAV: Journal of Information and Visualization*, 6(1). <https://doi.org/10.35877/454RI.jinav4161>
- Aziz, F., Ishak, P., & Abasa, S. (2024). Klasifikasi Depresi Menggunakan Support Vector Machine: Pendekatan Berbasis Data Text Mining. *Journal Pharmacy and Application of Computer Sciences*, 2(2), 33–38.
- Caron, R. M., Noel, K., Reed, R. N., Sibel, J., & Smith, H. J. (2024). Health Promotion, Health Protection, and Disease Prevention: Challenges and Opportunities in a Dynamic Landscape. *AJPM Focus*, 3(1). <https://doi.org/10.1016/j.focus.2023.100167>
- Cervantes, J., Garcia-Lamont, F., Rodríguez-Mazahua, L., & Lopez, A. (2020). A comprehensive survey on support vector machine classification: Applications, challenges and trends. *Neurocomputing*, 408, 189–215. <https://doi.org/https://doi.org/10.1016/j.neucom.2019.10.118>
- Chandel, A. (2025). Healthchare Chatbot Using SVM & Decision Tree. *Trends in health informatics*, 2(1), 10–17.

- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273–297. <https://doi.org/10.1007/BF00994018>
- DeCoster, J., Francis, M. K., Lee, G., & Rubinstein, E. (2026). Systematic Data Validation: Improving Research by Improving Data Quality. *Practical Assessment, Research, and Evaluation*, 31(1). <https://doi.org/10.7275/PARE.3586>
- Deng, W. S., & Sloutsky, V. M. (2015). The development of categorization: Effects of classification and inference training on category representation. *Developmental Psychology*, 51(3), 392–405. <https://doi.org/10.1037/A0038749>
- Držík, D., & Kelebercová, L. (2025). Back-translation effects on static and contextual word embeddings for topic classification embedding in classification tasks. *PLOS One*, 20(8), e0330622. <https://doi.org/10.1371/journal.pone.0330622>
- Felix Liman, C., Sufinata, S., Irviantina, S., & Winardi, S. (2023). Implementation of BiLSTM-SVM Algorithm to Detect Fake News on Text-Based Media. *Jurnal CoreIT*, 9(2).
- França, R. P., Borges Monteiro, A. C., Arthur, R., & Iano, Y. (2021). An overview of deep learning in big data, image, and signal processing in the modern digital age. Dalam *Trends in Deep Learning Methodologies* (hlm. 63–87). Elsevier. <https://doi.org/10.1016/B978-0-12-822226-3.00003-9>
- GoodStats. (2023, Juni 7). *Konten Kesehatan Jadi yang Paling Banyak Diakses Masyarakat Indonesia 2023*. <https://data.goodstats.id/statistic/konten-kesehatan-jadi-yang-paling-banyak-diakses-masyarakat-indonesia-2023-2qYFO>
- Gunawan, T. S., Babiker, A. B. F., Ismail, N., & Effendi, M. R. (2021). Development of Intelligent Telegram Chatbot Using Natural Language Processing. *2021 7th International Conference on Wireless and Telematics (ICWT)*, 1–5. <https://doi.org/10.1109/ICWT52862.2021.9678471>
- Hasanati, N., Utami, T. S., & Kusumaningtyas, R. H. (2023). Sentiment Analysis on News Headlines of Nation's Capital Relocation Using CNN and SVM. *2023 11th International Conference on Cyber and IT Service Management (CITSM)*, 1–6.
- Karim, A. A. J., Asad, K. H. M., & Azam, A. (2025). *Strengthening False Information Propagation Detection: Leveraging SVM and Sophisticated Text Vectorization Techniques in comparison to BERT*. <http://arxiv.org/abs/2411.12703>

- Kementrian Komunikasi dan Digital. (2025, Januari 8). *Komdigi Identifikasi 1.923 Konten Hoaks Sepanjang Tahun 2024*. <https://www.komdigi.go.id/berita/siaran-pers/detail/komdigi-identifikasi-1923-konten-hoaks-sepanjang-tahun-2024>
- Kessler, S. H., & Bachmann, E. (2022). Debunking health myths on the internet: the persuasive effect of (visual) online communication. *Journal of Public Health*, 30(8), 1823–1835. <https://doi.org/10.1007/s10389-022-01694-3>
- Koesten, L., Simperl, E., Blount, T., Kacprzak, E., & Tennison, J. (2020). Everything you always wanted to know about a dataset: Studies in data summarisation. *International Journal of Human-Computer Studies*, 135, 102367. <https://doi.org/10.1016/j.ijhcs.2019.10.004>
- Kožuh, I., & Čakš, P. (2023). Social Media Fact-Checking: The Effects of News Literacy and News Trust on the Intent to Verify Health-Related Information. *Healthcare*, 11(20), 2796. <https://doi.org/10.3390/HEALTHCARE11202796>
- Laeli, A. R., Rustam, Z., Hartini, S., Maulidina, F., & Aurelia, J. E. (2020). Hyperparameter Optimization on Support Vector Machine using Grid Search for Classifying Thalassemia Data. *2020 International Conference on Decision Aid Sciences and Application (DASA)*, 817–821. <https://doi.org/10.1109/DASA51403.2020.9317227>
- Li, Q., Peng, H., Li, J., Xia, C., Yang, R., Sun, L., Yu, P. S., & He, L. (2022). A Survey on Text Classification: From Traditional to Deep Learning. *ACM Trans. Intell. Syst. Technol.*, 13(2). <https://doi.org/10.1145/3495162>
- Liu, H., Chen, X., & Liu, X. (2022). A Study of the Application of Weight Distributing Method Combining Sentiment Dictionary and TF-IDF for Text Sentiment Analysis. *IEEE Access*, 10, 32280–32289. <https://doi.org/10.1109/ACCESS.2022.3160172>
- Liu, R., Liu, E., Yang, J., Li, M., & Wang, F. (2006). Optimizing the Hyperparameters for SVM by Combining Evolution Strategies with a Grid Search. Dalam D.-S. Huang, K. Li, & G. W. Irwin (Ed.), *Intelligent Control and Automation: International Conference on Intelligent Computing, ICIC 2006 Kunming, China, August 16–19, 2006* (hlm. 712–721). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-37256-1_87
- Lopez, E., Gorla, G., Etxebarria-Elezgarai, J., Amigo, J. M., & Seifert, A. (2026). The importance of choosing a proper validation strategy in predictive models. Part 2: Recipes for (avoiding) overfitting-A tutorial. *Analytica Chimica Acta*, 1384, 344838. <https://doi.org/10.1016/J.ACA.2025.344838>

- Lubis, A. S., & Putri, R. A. (2025). Analisis Sentimen Mahasiswa Terhadap Penggunaan E-Learning dengan Algoritma Support Vector Machine (SVM). *Decode: Jurnal Pendidikan Teknologi Informasi*, 5(2), 778–788.
- Mahing, N. F., Gunawan, A. L., Zen, A. F. A., Bachtiar, F. A., & Wicaksono, S. A. (2024). Klasifikasi Tingkat Stress dari Data Berbentuk Teks dengan Menggunakan Algoritma Support Vector Machine (SVM) dan Random Forest. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 11(5), 1067–1076. <https://doi.org/10.25126/jtiik.2024118010>
- Martindale, H., Rowland, E., Flower, T., & Clews, G. (2020). Semi-supervised machine learning with word embedding for classification in price statistics. *Data & Policy*, 2, e12. <https://doi.org/10.1017/dap.2020.13>
- Mohammed, I. A., Kuta, I. I., Falode, O. C., & Bello, A. (2024). Comparative performance of undergraduate students in micro-teaching using Telegram and WhatsApp in collaborative learning settings. *Journal of Mathematics and Science Teacher*, 4(2), em063. <https://doi.org/10.29333/MATHSCITEACHER/14411>
- Mohammed, S., Budach, L., Feuerpfeil, M., Ihde, N., Nathansen, A., Noack, N., Patzlaff, H., Naumann, F., & Harmouch, H. (2025). The effects of data quality on machine learning performance on tabular data. *Information Systems*, 132, 102549. <https://doi.org/10.1016/j.is.2025.102549>
- Muhyidin, A., Setiawan, M. A. F., & Nurkhamid. (2021). Developing UNYSA Chatbot as Information Services about Yogyakarta State University. *Journal of Physics: Conference Series*, 1737(1), 012038. <https://doi.org/10.1088/1742-6596/1737/1/012038>
- Mumuni, A., & Mumuni, F. (2022). Data augmentation: A comprehensive survey of modern approaches. *Array*, 16, 100258. <https://doi.org/10.1016/j.array.2022.100258>
- Nainggolan, P. I., Prasvita, D. S., & Bukit, D. S. (2021). Klasifikasi Informasi Kesehatan Pada Data Media Sosial Menggunakan Support Vector Machine dan K-Fold Cross Validation. *Malikussaleh Journal of Mechanical Science and Technology*, 5(2), 34–38.
- Naufaldi, M., Jovita, S., Frans, C. M., Hanafi, M. L. I., & Qomariyah, N. N. (2021). COBY: COVID-19 Telegram Chatbot by Employing Machine Learning Algorithms. *2021 International Conference on ICT for Smart Society (ICISS)*, 1–7. <https://doi.org/10.1109/ICISS53185.2021.9533198>
- Nguyen, T., Nguyen, H. T., & Nguyen-Hoang, T. A. (2025). Data quality management in big data: Strategies, tools, and educational implications.

- Journal of Parallel and Distributed Computing*, 200, 105067. <https://doi.org/10.1016/J.JPDC.2025.105067>
- Papík, R. (2013). Introduction to Information Science. *Journal of Documentation*, 69(3), 474–476. <https://doi.org/10.1108/JD-04-2013-0049>
- Pratama, F. W., Hasan, K., Jannah, R., Zahara, E., & Syafitri, A. (2024). Conventional Mass Media Versus New Media. *Proceedings of International Conference on Social Science, Political Science, and Humanities (ICoSPOLHUM)*, 4, 12.
- Putri, R. K., & Athoillah, M. (2021). Support Vector Machine Untuk Identifikasi Berita Hoax Terkait Virus Corona (Covid-19). *Jurnal Informatika: Jurnal Pengembangan IT*, 6(3), 162–167.
- Reis, J. C. S., & Benevenuto, F. (2021). Supervised Learning for Misinformation Detection in WhatsApp. *Proceedings of the Brazilian Symposium on Multimedia and the Web, WebMedia '21*, 245–252. <https://doi.org/10.1145/3470482.3479641>
- Ridha, M., Abdur Rohman, M. K., Agustin, D., Shaputra, D. H., Manayla, Y., & Malikah, A. H. (2025). Penerapan Machine Learning untuk Klasifikasi Teks Depresi pada Kesehatan Mental dengan SVM, TF-IDF, dan Chi-Square. *Journal Software, Hardware and Information Technology*, 5(2), 171–182. <https://doi.org/10.24252/shift.v5i2.210>
- Ropikoh, I. A., Abdulhakim, R., Enri, U., & Sulistiyowati, N. (2021). Penerapan Algoritma Support Vector Machine (SVM) untuk Klasifikasi Berita Hoax Covid-19. *Journal of Applied Informatics and Computing*, 5(1), 64–73. <https://doi.org/10.30871/JAIC.V5I1.3167>
- Sani, R. R., Pratiwi, Y. A., Winarno, S., Udayanti, E. D., & Alzami, F. (2022). Analisis Perbandingan Algoritma Naive Bayes Classifier dan Support Vector Machine untuk Klasifikasi Berita Hoax pada Berita Online Indonesia. *Jurnal Masyarakat Informatika*, 13(2), 85–98.
- Schlicht, I. B., Fernandez, E., Chulvi, B., & Rosso, P. (2024). Automatic detection of health misinformation: a systematic review. *Journal of Ambient Intelligence and Humanized Computing*, 15(3), 2009–2021.
- ShamsEldin, T., Gaber, S., Ansari, S., Elgohary, R., Shawky, M. A., Elbahnasawy, M., & Abdrabou, M. (2025). Artificial intelligence for predicting depression anxiety and stress using psychometric data. *Scientific Reports*, 15(1), 37282. <https://doi.org/10.1038/s41598-025-21301-1>
- Sikosana, M., Ajao, O., & Maudsley-Barton, S. (2024). A Comparative Study of Hybrid Models in Health Misinformation Text Classification. *OASIS 2024 -*

Proceedings of the 2024 Workshop on Open Challenges in Online Social Media, Held in conjunction with the 35th ACM Conference on Hypertext and Social Media, HT 2024, 18–25. <https://doi.org/10.1145/3677117.3685007>

- Sitepu, R., Rohmawati, A. A., & Prasetyowati, S. S. (2022). The Analysis of Support Vector Machine (SVM) on Monthly Covid-19 Case Classification. *International Journal on Information and Communication Technology (IJoICT)*, 8(2), 40–52. <https://doi.org/10.21108/IJOICT.V8I2.671>
- Sitohang, D., & Lestari, M. W. (2025). Perspektif Mahasiswa Teknologi Laboratorium Medis Terhadap Kasus Penyebaran Informasi Hoax Bidang Kesehatan Melalui Media Sosial: Perceptions of Medical Laboratory Technology Students at the National College of Health Sciences regarding Cases of Dissemination of Hoax Information in the Health Sector via SocialMedia. *Journal of Nursing and Health*, 10(1), 38–49.
- Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, 45(4), 427–437. <https://doi.org/https://doi.org/10.1016/j.ipm.2009.03.002>
- Sudhakar, M., & Kaliyamurthi, K. P. (2024). Detection of fake news from social media using support vector machine learning algorithms. *Measurement: Sensors*, 32, 101028. <https://doi.org/10.1016/J.MEASEN.2024.101028>
- Supriyono, Wibawa, A. P., Suyono, & Kurniawan, F. (2024). Advancements in natural language processing: Implications, challenges, and future directions. *Telematics and Informatics Reports*, 16, 100173. <https://doi.org/https://doi.org/10.1016/j.teler.2024.100173>
- Syarah, M. M., Sarwoprasodjo, S., & Lumintang, R. W. E. (2014). The Role of Health Communication among the Poor. *Makara Human Behavior Studies in Asia*, 18(2), 149–158. <https://doi.org/10.7454/mssh.v18i2.3469>
- T R, M., V, V. K., V, D. K., Geman, O., Margala, M., & Guduri, M. (2023). The stratified K-folds cross-validation and class-balancing methods with high-performance ensemble classifiers for breast cancer classification. *Healthcare Analytics*, 4, 100247. <https://doi.org/https://doi.org/10.1016/j.health.2023.100247>
- Tajrian, M., Rahman, A., Kabir, M. A., & Islam, M. R. (2024). Analysis of child development facts and myths using text mining techniques and classification models. *Heliyon*, 10(17), e36652. <https://doi.org/https://doi.org/10.1016/j.heliyon.2024.e36652>
- Wei, Z., & Zhang, S. (2024). A structured sentiment analysis dataset based on public comments from various domains. *Data in Brief*, 53, 110232. <https://doi.org/https://doi.org/10.1016/j.dib.2024.110232>

- Widyanto, T., Ristiana, I., & Wibowo, A. (2023). Komparasi Naïve Bayes dan SVM Analisis Sentimen RUU Kesehatan di Twitter. *SINTECH (Science and Information Technology) Journal*, 6(3), 147–161.
- Woodhouse, C. (2023). Covid myths persist despite evidence to refute them. *Trends in Urology & Men's Health*, 14(6), 2–6. <https://doi.org/https://doi.org/10.1002/tre.939>
- Yildirim-Erbasli, S. N., & Bulut, O. (2026). Exploring Data Augmentation in a Low-Resource Language Context: A Case Study on Text Generation for Reading Comprehension in Turkish. *Algorithms*, 19(5), 413. <https://doi.org/10.3390/a19050413>
- Yu, J., Bekerian, D. A., & Osback, C. (2024). Navigating the Digital Landscape: Challenges and Barriers to Effective Information Use on the Internet. *Encyclopedia*, 4(4), 1665–1680. <https://doi.org/10.3390/encyclopedia4040109>