pISSN: 2442-3386 eISSN: 2442-4293
Melek IT
Vol 11 No 1 Jan – Jun 2025, 101 - 112
INFORMATION TECHNOLOGY JOURNAL

PENGEMBANGAN SISTEM DETEKSI PHISHING PADA WHATSAPP BERBASIS AI DENGAN INTEGRASI BAHASA LOKAL SURABAYA

Fahmi Mochtar Efendi[1], Anang Kukuh Adisusilo *[2] , Slamet Budiprayitno[3]
[1]Program Studi Informatika, Universitas Wijaya Kusuma Surabaya, fahmiefendi644@gmail.com
[2]Program Studi Informatika, Universitas Wijaya Kusuma Surabaya, anang65@uwks.ac.id
[3]Departemen Elektro Otomasi, Institut Teknologi Sepuluh Nopember (ITS), slametbp1378@its.ac.id
*)Korespondensi: anang65@uwks.ac.id

Abstrak

Phishing merupakan salah satu bentuk kejahatan siber yang kian meningkat seiring dengan tingginya penggunaan teknologi komunikasi digital, khususnya melalui platform pesan instan seperti WhatsApp. Modus operandi phishing sering kali memanfaatkan pesan yang berisi tautan palsu dan menggunakan bahasa yang meyakinkan untuk mengecoh dan memanipulasi psikologis korban agar mengakses tautan atau memberikan informasi pribadi. Dalam rangka memberikan solusi terhadap permasalahan tersebut, penelitian ini bertujuan untuk merancang dan mengimplementasikan sebuah sistem pendeteksi phishing berbasis kecerdasan buatan (Artificial Intelligence/AI), dengan mengadopsi dua pendekatan logika fuzzy, yaitu Fuzzy Mamdani dan Fuzzy Takagi-Sugeno-Kang (TSK). Sistem ini diintegrasikan ke dalam platform WhatsApp dengan memanfaatkan parameter-parameter utama seperti jumlah kata kunci yang mencurigakan, keberadaan tautan dalam pesan, dan panjang karakter pesan. Selain itu, pendekatan yang digunakan mempertimbangkan konteks bahasa lokal yang sering digunakan oleh pengguna, termasuk Bahasa Indonesia, Bahasa Jawa, serta dialek khas daerah seperti dialek Surabaya, sehingga sistem lebih adaptif terhadap variasi linguistik lokal. Dalam implementasinya, model AI dikembangkan menggunakan framework FastAPI dan dikoneksikan dengan bot WhatsApp berbasis Node.js untuk mendukung komunikasi secara real-time. Hasil pengujian menunjukkan bahwa kedua pendekatan fuzzy yang digunakan mampu mendeteksi pesan phishing secara tepat dengan nilai tingkat keyakinan (confidence) yang stabil dan relevan, baik melalui perhitungan manual maupun hasil aktual dari sistem. Sistem ini juga memberikan peringatan otomatis kepada pengguna maupun dalam grup WhatsApp, serta merekam hasil deteksi ke dalam database untuk keperluan analisis lanjutan. Temuan ini menunjukkan potensi besar sistem dalam meningkatkan keamanan komunikasi digital berbasis pesan instan.

Kata Kunci: Phishing, Artificial Intelligence, Fuzzy Logic, WhatsApp Bot, Deteksi Pesan Berbahaya.

Abstract

Phishing is a form of cybercrime that continues to grow, especially through instant messaging platforms such as WhatsApp. Phishing messages often use fake links and persuasive language to deceive victims. This study aims to design and implement a phishing detection system based on Artificial Intelligence (AI) using two fuzzy logic approaches: Fuzzy Mamdani and Fuzzy Takagi-Sugeno-Kang (TSK), integrated into the WhatsApp platform. The system detects phishing based on several parameters, including the number of suspicious keywords, the presence of URLs, and message length, while also considering the local language context such as Indonesian, Javanese, and the Surabaya dialect. The AI model was developed using the FastAPI framework and connected to a WhatsApp bot built with Node.js for real-time communication. Implementation results show that both fuzzy models can accurately detect phishing messages in real time, with confidence levels that are consistent between manual calculations and system outputs. The system also provides automatic alerts to individual users and group chats and stores detection results in a database for further analysis. These findings highlight the effectiveness of AI-based detection systems in enhancing digital communication security, especially when adapted to regional language characteristics. [2], [5], [12]

Keywords : Phishing, Artificial Intelligence, Fuzzy Logic, WhatsApp Bot, Local Language Detection

I. PENDAHULUAN

Phishing merupakan salah satu bentuk serangan siber yang paling sering terjadi dan memiliki dampak signifikan terhadap keamanan data pengguna. Serangan ini biasanya dilakukan dengan mengirimkan pesan palsu yang tampak seolah-olah berasal dari pihak terpercaya, dengan tujuan untuk mencuri informasi sensitif seperti kata sandi, data pribadi, atau kredensial akun. Seiring meningkatnya penggunaan aplikasi pesan instan sebagai media komunikasi utama, platform seperti WhatsApp menjadi sasaran empuk bagi pelaku phishing. Menurut laporan dari Kaspersky [1], jumlah insiden phishing yang menyasar aplikasi perpesanan

meningkat drastis sepanjang tahun 2022. Hal ini didukung oleh data dari We Are Social [2] yang menunjukkan bahwa WhatsApp merupakan aplikasi perpesanan dengan jumlah pengguna terbanyak secara global, termasuk di Indonesia. [1], [3], [6]

Meskipun berbagai upaya telah dilakukan untuk mengurangi penyebaran pesan phishing, pendekatan konvensional seperti filter kata kunci atau deteksi berbasis aturan tetap memiliki keterbatasan. Metode tersebut kurang mampu mengenali variasi bahasa yang digunakan dalam pesan, terutama dalam konteks lokal yang bercampur antara Bahasa Indonesia, Jawa, atau dialek Surabaya. Selain itu, sebagian besar sistem deteksi phishing saat ini belum terintegrasi langsung dengan platform pesan instan seperti WhatsApp, sehingga tidak dapat memberikan respon secara real-time kepada pengguna. [4], [11], [13]

Penelitian sebelumnya menunjukkan bahwa pendekatan logika fuzzy efektif dalam menangani ketidakpastian dan ambiguitas dalam pesan phishing. Studi oleh Tarek et al. menunjukkan bahwa model berbasis fuzzy logic dapat mendeteksi email phishing dengan akurasi tinggi dan keputusan yang lebih fleksibel [6]. Selain itu, pendekatan Takagi-Sugeno-Kang (TSK) fuzzy yang dioptimalkan dengan algoritma gradient descent terbukti mampu meningkatkan akurasi sistem hingga 99,95% serta menjaga interpretabilitas aturan yang tinggi dalam sistem keamanan [7]. Pendekatan fuzzy lain, seperti Fuzzy Rule Interpolation (FRI), juga menunjukkan keunggulan dalam mendeteksi situs phishing dengan tingkat akurasi 97,6% dan false positive rate yang rendah [9]. Hal ini memperkuat dasar bahwa logika fuzzy dengan berbagai pendekatannya layak diadopsi untuk kasus phishing pada pesan instan. [1], [3], [6]

Dari sisi lokal, penelitian oleh Prasetyo dan Adisusilo menunjukkan penerapan fuzzy logic dalam sistem pendukung keputusan berbasis risiko transaksi daring, yang meskipun konteksnya berbeda, prinsip desain dan evaluasi fuzzy yang digunakan sangat relevan untuk dikembangkan dalam deteksi pesan phishing berbasis konteks lokal seperti dialek Surabaya [8]. [14], [15]

Kondisi ini menciptakan kebutuhan mendesak akan sistem cerdas yang mampu mengenali potensi phishing secara otomatis dengan mempertimbangkan konteks linguistik dan karakteristik pesan pengguna secara fleksibel.

Untuk menjawab permasalahan tersebut, penelitian ini mengusulkan pembangunan sistem pendeteksian phishing berbasis kecerdasan buatan (AI) yang diintegrasikan langsung ke platform WhatsApp. Sistem ini menggunakan dua pendekatan fuzzy logic, yaitu Fuzzy Mamdani dan Fuzzy Takagi-Sugeno-Kang (TSK), untuk mengevaluasi potensi phishing berdasarkan parameter seperti jumlah kata kunci mencurigakan, keberadaan tautan, dan panjang pesan. Keunikan sistem ini terletak pada kemampuannya mendeteksi bahasa lokal seperti Bahasa Indonesia, Jawa, dan dialek Surabaya, sehingga lebih relevan dengan konteks penggunaan di komunitas kampus atau daerah. Sistem ini diimplementasikan menggunakan kombinasi FastAPI sebagai backend AI dan bot WhatsApp berbasis Node.js untuk interaksi langsung dengan pengguna. [2], [5], [12].

Meskipun berbagai pendekatan telah dilakukan dalam mendeteksi phishing menggunakan teknik kecerdasan buatan, sebagian besar studi masih berfokus pada data berbasis surel (email) atau website, dan belum banyak yang mengkaji secara khusus pesan teks pada aplikasi pesan instan seperti *WhatsApp*. Di sisi lain, penelitian yang mengadopsi pendekatan *fuzzy logic* sering kali hanya memanfaatkan bahasa standar seperti Bahasa Indonesia formal atau Bahasa Inggris, tanpa mempertimbangkan keberagaman konteks linguistik lokal yang umum digunakan dalam komunikasi sehari-hari.

Kesenjangan ini menunjukkan bahwa belum banyak sistem yang dirancang untuk menangani karakteristik pesan phishing dalam konteks budaya lokal, seperti penggunaan Bahasa Jawa atau dialek Surabaya. Selain itu, sangat sedikit penelitian yang mengintegrasikan langsung sistem deteksi phishing berbasis fuzzy logic ke dalam bot WhatsApp yang mampu memberikan respon otomatis secara real-time kepada pengguna.

Berdasarkan celah tersebut, penelitian ini berkontribusi secara ilmiah dalam tiga aspek utama. Pertama, menggabungkan dua pendekatan fuzzy yaitu Mamdani dan Takagi-Sugeno-Kang (TSK) dalam mendeteksi phishing berbasis pesan teks pendek. Kedua, menyesuaikan sistem dengan konteks bahasa lokal seperti Bahasa Indonesia, Jawa, dan Surabaya untuk meningkatkan relevansi deteksi terhadap komunitas lokal. Ketiga, mengintegrasikan model ini langsung ke dalam platform WhatsApp melalui FastAPI dan Node.js, sehingga memungkinkan pengujian dan penerapan secara praktis dalam komunikasi sehari-hari.

Penelitian ini bertujuan untuk: (1) merancang dan mengimplementasikan sistem deteksi phishing otomatis berbasis fuzzy logic di platform WhatsApp; (2) menerapkan dan membandingkan dua metode fuzzy, yaitu Mamdani dan TSK, dalam konteks klasifikasi pesan phishing; dan (3) mengevaluasi efektivitas sistem berdasarkan hasil perhitungan manual dan implementasi aktual. Lingkup sistem dibatasi pada analisis pesan berbasis teks, dengan fokus pada parameter kata kunci mencurigakan, panjang pesan, dan keberadaan tautan, tanpa mencakup media seperti gambar atau file. Artikel ini disusun dalam beberapa bagian, yaitu: bagian metode yang menjelaskan desain sistem dan model fuzzy; bagian hasil dan pembahasan yang memaparkan implementasi dan visualisasi; serta bagian penutup yang merangkum kesimpulan dan saran. [4], [11], [13]

pISSN: 2442-3386 eISSN: 2442-4293
Vol 11 No 1 Jan – Jun 2025, 103 - 112
Melek IT
INFORMATION TECHNOLOGY JOURNAL

II. METODE

2.1 Desain Sistem

Sistem yang dikembangkan bertujuan untuk mendeteksi pesan phishing secara otomatis pada platform WhatsApp dengan memanfaatkan metode kecerdasan buatan berbasis fuzzy logic. Pendekatan fuzzy logic, terutama fuzzy interval tipe2, terbukti efektif dalam mendeteksi spam di media sosial dengan kombinasi fitur yang kompleks [3]. Sistem ini dirancang untuk memberikan peringatan secara real-time kepada pengguna dan grup jika ditemukan pesan yang mengandung indikasi phishing. Peringatan tersebut bertujuan untuk mencegah penyebaran pesan berbahaya yang dapat membahayakan data atau informasi pengguna. Desain agent AI dalam sistem ini mengikuti konsep perception–reasoning–action sebagaimana dijelaskan oleh Russell dan Norvig [4]. Sistem ini juga mempertimbangkan konteks lokal dalam analisis, seperti penggunaan Bahasa Indonesia, Jawa, dan dialek Surabaya, guna meningkatkan akurasi deteksi pada komunitas tertentu. [1], [3], [6]

Secara arsitektur, sistem terdiri dari tiga komponen utama yang saling terintegrasi. Pertama, bot WhatsApp yang bertugas menerima dan mengirim pesan melalui library whatsapp-web.js. Kedua, backend AI berbasis FastAPI yang menjalankan dua model fuzzy (Mamdani dan TSK) untuk mengevaluasi potensi phishing dari pesan masuk. Ketiga, database MySQL yang digunakan untuk menyimpan hasil deteksi lengkap, termasuk isi pesan, confidence score, status phishing, bahasa terdeteksi, dan metadata lainnya. Komunikasi antar komponen berjalan secara otomatis: pesan dari pengguna diteruskan oleh bot ke API, diproses oleh model AI, kemudian hasilnya dikembalikan ke bot untuk dikirimkan sebagai balasan dan dicatat dalam database. [2], [5], [12]

Alur kerja sistem dimulai saat pengguna mengirimkan pesan ke grup atau kontak WhatsApp yang terhubung dengan bot. Bot akan membaca isi pesan dan meneruskannya ke server backend melalui permintaan HTTP. Di sisi backend, sistem pertama-tama melakukan deteksi bahasa untuk mengidentifikasi apakah pesan ditulis dalam Bahasa Indonesia, Jawa, atau dialek Surabaya. Selanjutnya, sistem melakukan ekstraksi fitur dari pesan, seperti jumlah kata kunci mencurigakan, keberadaan tautan, dan panjang pesan. Nilai-nilai ini kemudian difuzzifikasi dan diproses oleh dua model fuzzy logic, yaitu Mamdani dan TSK, untuk menghasilkan confidence score yang menunjukkan tingkat kemungkinan pesan tersebut tergolong phishing. [4], [11], [13]

Sebagai hasil dari pemrosesan tersebut, sistem akan memberikan balasan kepada pengirim pesan berupa peringatan jika *confidence score* melebihi ambang batas yang telah ditentukan. Selain itu, sistem juga mengirimkan notifikasi otomatis ke grup sebagai tindakan mitigasi awal, dan mencatat seluruh hasil analisis ke dalam *database* untuk keperluan audit dan evaluasi. Dengan desain ini, sistem mampu bekerja secara *real-time* tanpa campur tangan manual, serta adaptif terhadap variasi bahasa lokal yang digunakan dalam pesan sehari-hari.

2.2 Parameter dan Data Input

Untuk mengevaluasi apakah sebuah pesan tergolong *phishing* atau tidak, sistem menggunakan tiga parameter utama yang diekstraksi langsung dari isi pesan teks. Ketiga parameter tersebut adalah: (1) jumlah kata kunci mencurigakan, (2) keberadaan tautan atau *link*, dan (3) panjang pesan dalam satuan karakter. Kata kunci mencurigakan mengacu pada daftar kata yang umum digunakan dalam pesan *phishing* seperti "verifikasi", "hadiah", "klik", "token", dan sejenisnya. Sistem akan menghitung berapa banyak kata dari daftar tersebut yang muncul dalam pesan, yang kemudian diklasifikasikan sebagai *few*, *moderate*, atau *many* dalam proses fuzzifikasi.

Parameter kedua, yaitu keberadaan tautan, diperiksa dengan mendeteksi apakah pesan mengandung pola *URL* (seperti "http://", "https://", atau domain umum). Parameter ini bersifat biner: jika ada link maka bernilai 1 (*yes*), jika tidak maka 0 (*no*). Parameter ketiga adalah panjang pesan, yang diukur berdasarkan jumlah karakter dalam pesan. Panjang pesan dibagi menjadi tiga kategori *fuzzy*, yaitu *short*, *medium*, dan *long*. Selain ketiga parameter utama tersebut, sistem juga mendeteksi bahasa pesan menggunakan *library langdetect*, yang akan mengklasifikasikan pesan ke dalam Bahasa Indonesia, Jawa, atau dialek Surabaya. Hasil deteksi bahasa ini dicatat dalam *database* sebagai bagian dari konteks pesan, meskipun tidak digunakan langsung dalam proses *fuzzy*.

2.3 Fuzzifikasi dan *Rule Base*

Setelah parameter numerik diperoleh dari isi pesan, tahap selanjutnya adalah proses *fuzzifikasi*, yaitu konversi nilai-nilai tersebut ke dalam bentuk variabel linguistik yang dapat diproses oleh sistem *fuzzy*. Untuk parameter jumlah kata kunci mencurigakan, digunakan tiga fungsi keanggotaan segitiga: *few* (0–3), *moderate* (2–6), dan *many* (5–10). Panjang pesan juga difuzzifikasi ke dalam tiga kategori: *short* (0–100 karakter), *medium* (80–200 karakter), dan *long* (180–300 karakter). Sementara itu, parameter keberadaan tautan menggunakan representasi *biner* karena bersifat eksplisit: 0 untuk *no* dan 1 untuk *yes*.

Sistem menggunakan dua model *fuzzy*, yaitu *Fuzzy Mamdani* dan *Fuzzy TSK*. Pada model Mamdani, digunakan aturan linguistik *IF–THEN* seperti: "IF jumlah kata kunci adalah many AND link adalah yes THEN output adalah high." Hasil dari semua aturan *Mamdani* digabungkan dan didefuzzifikasi menggunakan metode *centroid* untuk menghasilkan

nilai *confidence* antara 0 hingga 1. Sedangkan pada model *TSK*, setiap aturan memiliki output berupa fungsi linier dari *input* numerik, dan hasil akhir diperoleh melalui perhitungan rata-rata berbobot (*weighted average*). Kedua model ini digunakan secara paralel untuk menilai tingkat risiko *phishing* dari setiap pesan yang dianalisis. Sistem deteksi *phishing* berbasis *e-banking* menggunakan metode *fuzzy data mining* telah menunjukkan efektivitas dalam menangani ambiguitas pada atribut *phishing* seperti *URL* dan *domain* [5].

2.4 Perhitungan Manual

Sebagai verifikasi terhadap hasil implementasi sistem, dilakukan perhitungan manual pada satu kasus pesan uji untuk menilai akurasi model *fuzzy Mamdani* dan *TSK*. Contoh pesan yang digunakan adalah:

"Cek link ini untuk klaim hadiah kamu
https://bit.ly/claim"

Dari pesan tersebut diperoleh tiga parameter input: jumlah kata kunci mencurigakan sebanyak 3 (*cek*, *klaim*, *hadiah*), panjang pesan 95 karakter, dan keberadaan tautan (*ya*). Nilai-nilai ini kemudian difuzzifikasi seperti pada Tabel 1.

Tabel 1. Nilai-nilai difuzzifikasi

Parameter	Nilai	*Fuzzy Set*	Derajat Keanggotaan
Kata Kunci Mencurigakan	3	*Few, Moderate*	0.5, 0.5
Panjang Pesan	95	*Short, Medium*	0.25, 0.75
Link	Ya	*Yes*	1.0

Pada model *Mamdani*, digunakan aturan seperti:

a) *IF Moderate AND Medium AND Yes THEN High*
b) *IF Few AND Medium AND Yes THEN Medium*

Evaluasi minimum derajat keanggotaan menghasilkan:

a) *Rule* untuk *Medium*: min(0.5, 0.75, 1.0) = 0.5
b) *Rule* untuk *High*: min(0.5, 0.75, 1.0) = 0.5

Hasil defuzzifikasi *centroid* dari kombinasi kedua *rule* tersebut menghasilkan *confidence* sebesar 0.65. Sementara pada model *TSK*, dua *rule* aktif memberikan *output* linier masing-masing 0.4 dan 0.7 dengan bobot sama (0.5), menghasilkan *confidence* akhir 0.55.

Dengan hasil di atas, baik model *Mamdani* maupun *TSK* sama-sama mengklasifikasikan pesan tersebut sebagai *phishing*, karena nilai *confidence* melebihi ambang batas 0.5. Hasil ini konsisten dengan hasil implementasi sistem.

2.5 Implementasi Sistem

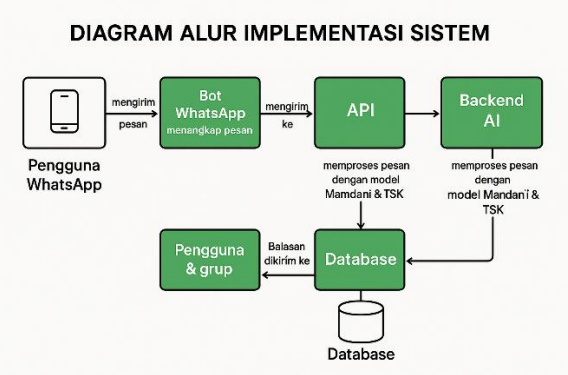
Sistem pendeteksian phishing ini diimplementasikan dengan pendekatan arsitektur terdistribusi yang menggabungkan beberapa teknologi utama. Bagian frontend interaksi pengguna berupa bot WhatsApp dikembangkan menggunakan Node.js dengan library whatsapp-web.js, yang memungkinkan komunikasi langsung dengan pengguna di platform WhatsApp. Backend atau server AI dibangun dengan menggunakan framework FastAPI berbasis Python, yang berfungsi untuk memproses pesan masuk dengan metode fuzzy logic Mamdani dan TSK. Untuk penyimpanan hasil deteksi, digunakan basis data MySQL yang menyimpan informasi isi pesan, confidence score dari kedua model, status phishing, bahasa pesan, serta metadata pengirim. [2], [5], [12]

Integrasi antar komponen dilakukan melalui komunikasi berbasis HTTP. Setiap kali pengguna mengirimkan pesan ke bot WhatsApp, pesan tersebut ditangkap oleh bot dan diteruskan ke endpoint API backend menggunakan metode POST. Backend kemudian melakukan deteksi bahasa dan ekstraksi fitur seperti jumlah kata kunci mencurigakan, keberadaan tautan, dan panjang pesan. Nilai-nilai ini kemudian difuzzifikasi dan diproses oleh model fuzzy Mamdani dan TSK secara paralel untuk menghasilkan nilai confidence. [4], [11], [13]

Setelah proses deteksi selesai, sistem akan mengembalikan hasil analisis ke bot WhatsApp dalam format JSON. Bot kemudian menampilkan balasan kepada pengirim pesan yang menyatakan status aman atau phishing, berdasarkan nilai confidence dari model Mamdani. Jika terindikasi phishing, maka bot juga mengirimkan notifikasi peringatan ke grup WhatsApp yang relevan. Secara bersamaan, hasil deteksi dicatat ke dalam tabel detections di MySQL untuk keperluan audit, evaluasi model, atau pelaporan keamanan. [4], [11], [13]

Dengan implementasi ini, sistem dapat bekerja secara otomatis dan *real-time* tanpa campur tangan manual. Kombinasi antara metode *fuzzy logic* dan pendekatan berbasis bahasa lokal membuat sistem adaptif terhadap variasi konteks pesan. Selain itu, desain modular sistem memungkinkan perluasan di masa depan, seperti penambahan model lain atau integrasi ke platform pesan instan lainnya.

2.6 Diagram Alur Implementasi

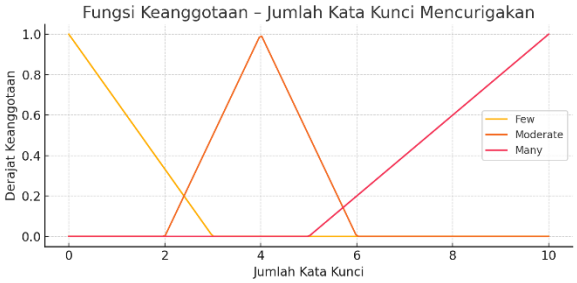


Gambar 1. Diagram Alur Implementasi Sistem

pISSN: 2442-3386 eISSN: 2442-4293
Vol 11 No 1 Jan – Jun 2025, 105 - 112
Melek IT
INFORMATION TECHNOLOGY JOURNAL

Gambar 1 menunjukkan alur implementasi sistem pendeteksian phishing yang dibangun. Proses dimulai dari pesan WhatsApp yang dikirimkan oleh pengguna, kemudian ditangkap oleh bot yang dibangun menggunakan Node.js. Pesan tersebut diteruskan melalui permintaan HTTP ke server backend berbasis FastAPI yang menjalankan dua model fuzzy (Mamdani dan TSK). Setelah dilakukan proses deteksi, hasil berupa nilai confidence dan status phishing dikembalikan ke bot untuk dikirimkan ke pengirim maupun grup admin. Di saat yang sama, hasil deteksi dicatat secara otomatis ke dalam database MySQL. Diagram ini membantu menjelaskan bagaimana integrasi antar komponen dilakukan secara real-time dan tanpa interaksi manual. [4], [11], [13]

2.7 Diagram Keanggotaan



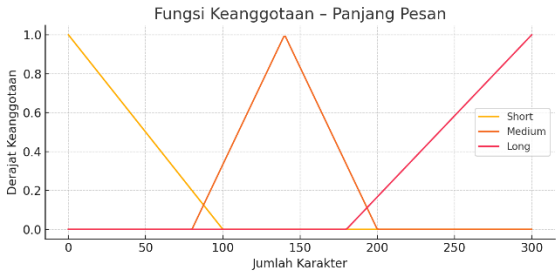
Gambar 2. Fungsi Keanggotaan – Jumlah Kata Kunci Mencurigakan

Gambar 2 menunjukkan fungsi keanggotaan *fuzzy* untuk parameter jumlah kata kunci mencurigakan. Fungsi ini digunakan untuk memetakan nilai numerik (jumlah kata kunci yang ditemukan dalam pesan) ke dalam kategori linguistik yang digunakan dalam aturan *fuzzy*, yaitu *few*, *moderate*, dan *many*.

1) *Few* (sedikit) diwakili oleh fungsi *linear* menurun, dengan derajat keanggotaan tertinggi saat jumlah kata kunci mendekati nol, dan turun hingga nol pada nilai 3.
2) *Moderate* (sedang) mempunyai bentuk segitiga dengan puncak di sekitar nilai 4. Rentang efektifnya berada antara 2 hingga 6, mencerminkan kondisi ambigu antara sedikit dan banyak.
3) *Many* (banyak) mulai meningkat setelah nilai 5 dan mencapai derajat keanggotaan penuh di angka 10.

Fungsi-fungsi ini memungkinkan sistem untuk menginterpretasikan jumlah kata kunci secara fleksibel dan menangani ketidakpastian dalam klasifikasi. Sebuah pesan dengan 4 kata kunci, misalnya, bisa memiliki derajat keanggotaan sebagian pada *moderate* dan *many*, tergantung bentuk fungsinya.

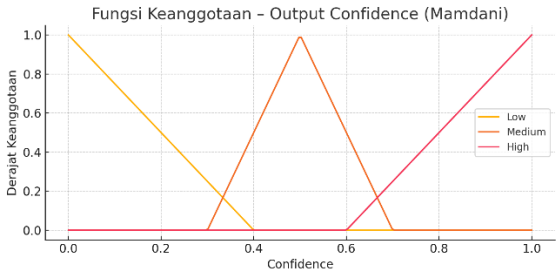
Grafik pada Gambar 3 menunjukkan bagaimana sistem fuzzy memetakan nilai panjang pesan (dalam karakter) ke dalam tiga kategori linguistik: *short*, *medium*, dan *long*. Klasifikasi ini digunakan dalam model *fuzzy* untuk mengevaluasi seberapa signifikan isi pesan terhadap potensi *phishing*.



Gambar 3. Fungsi Keanggotaan – Panjang Pesan

1) *Short*. Pesan dengan panjang mendekati 0 hingga sekitar 100 karakter memiliki derajat keanggotaan tinggi dalam kategori *short*. Fungsi ini menurun secara *linear* dan menjadi nol setelah 100 karakter.
2) *Medium*. Berbentuk segitiga dengan puncak di sekitar 140 karakter, merepresentasikan pesan berukuran sedang. Rentang efektifnya berada antara 80 hingga 200 karakter.
3) *Long*. Naik dari sekitar 180 karakter dan mencapai derajat keanggotaan penuh di 300 karakter. Mewakili pesan yang lebih panjang dan berpotensi mengandung banyak elemen mencurigakan.

Dengan pendekatan ini, sistem dapat mengenali bahwa panjang pesan dapat berkontribusi terhadap klasifikasi phishing, terutama jika digabungkan dengan jumlah kata kunci atau keberadaan link. Misalnya, pesan pendek biasanya dianggap tidak berbahaya, sedangkan pesan panjang dengan banyak kata kunci mencurigakan akan meningkatkan nilai confidence dalam model fuzzy. [1], [3], [6]



Gambar 4. Fungsi Keanggotaan – *Output Confidence* (*Mamdani*)

Grafik pada Gambar 4 menunjukkan fungsi keanggotaan yang digunakan untuk memetakan hasil akhir deteksi *phishing* dalam model *Fuzzy Mamdani*. *Output* ini merepresentasikan *confidence* atau tingkat keyakinan bahwa suatu pesan tergolong *phishing*, dengan rentang nilai antara 0 hingga 1. Terdapat tiga kategori linguistik: *low*, *medium*, dan *high*.

1) *Low*. *Confidence* dianggap rendah jika nilainya berada di bawah 0.4, dengan derajat keanggotaan tertinggi saat *confidence* mendekati nol.
2) *Medium*. Mencapai puncak pada nilai sekitar 0.5. Kategori ini menangkap ketidakpastian atau ambiguitas, dengan rentang dari 0.3 hingga 0.7.

3) *High*. *Confidence* dianggap tinggi jika nilainya berada di atas 0.6 dan mencapai maksimum di angka 1.

Sistem menggunakan nilai-nilai ini untuk memutuskan apakah suatu pesan akan diklasifikasikan sebagai *phishing* atau tidak. Jika hasil defuzzifikasi (dengan metode *centroid*) menghasilkan nilai lebih dari 0.5, maka pesan dikategorikan *phishing*, dan sistem akan mengirimkan peringatan. Fungsi keanggotaan ini memungkinkan transisi yang halus dan fleksibel antara level risiko yang berbeda.

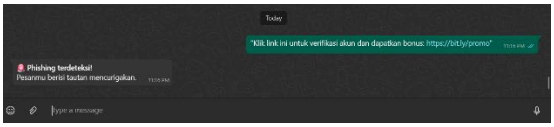
III. HASIL DAN PEMBAHASAN

3.1 Uji Coba Sistem

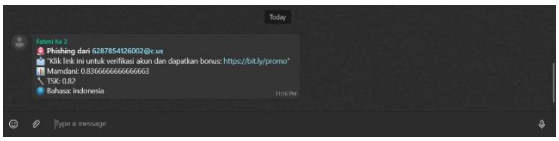
Untuk mengevaluasi performa sistem pendeteksian phishing yang dibangun, dilakukan serangkaian uji coba menggunakan skenario pengiriman pesan melalui platform WhatsApp. Tujuan dari pengujian ini adalah untuk menilai kemampuan sistem dalam mengenali pesan yang tergolong phishing maupun pesan yang bersifat aman. Sistem diuji dalam lingkungan simulasi yang merepresentasikan kondisi nyata, termasuk penggunaan bahasa informal, keberadaan tautan aktif, serta konteks pesan khas pengguna WhatsApp di lingkungan kampus. [4], [11], [13]

Pengujian dilakukan dengan menggunakan lima pesan uji yang terdiri dari kombinasi antara pesan phishing dan pesan non-phishing. Pesan-pesan tersebut dirancang mencerminkan karakteristik umum pesan WhatsApp, seperti ajakan mengklik tautan, informasi hadiah, atau instruksi tugas. Setiap pesan yang diterima bot WhatsApp diteruskan ke server backend FastAPI, lalu dianalisis secara paralel menggunakan dua metode, yaitu Fuzzy Mamdani dan Fuzzy TSK. Hasil dari kedua model berupa nilai confidence dan status klasifikasi (phishing atau aman) ditampilkan kembali ke pengirim dan dicatat dalam database MySQL. [4], [11], [13]

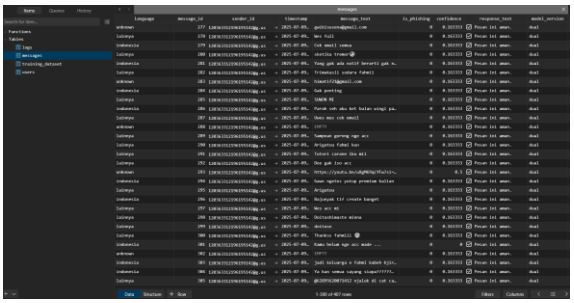
Selain memberikan klasifikasi terhadap isi pesan, sistem juga secara otomatis memberikan respon kepada pengguna dan grup admin. Jika pesan terindikasi phishing oleh model Mamdani, sistem mengirimkan peringatan berupa pesan balasan ke pengirim dan notifikasi ke grup WhatsApp. Sementara itu, semua hasil deteksi—baik dari Mamdani maupun TSK—dicatat ke dalam basis data secara lengkap, termasuk confidence score, status, dan metadata pesan. Proses ini berjalan secara real-time tanpa intervensi manual, sehingga sistem dapat digunakan untuk memitigasi penyebaran phishing secara cepat dan efisien. [4], [11], [13]



Gambar 5. Belasan Ke Pengirim Whatsapp [4], [11], [13]



Gambar 6. Notifikasi Otomatis Ke Grup



Gambar 7. Isi *Database Message* Deteksi

Tampilan pada Gambar 5 menunjukkan bagaimana sistem memberikan balasan otomatis kepada pengguna setelah mendeteksi pesan *phishing*. Gambar 6 menunjukkan sistem mengirimkan notifikasi ke grup *admin* berisi detail hasil deteksi. Sementara itu, Gambar 7 memperlihatkan *data* yang tercatat di *database*, termasuk *confidence score* dan status deteksi masing-masing model.

3.2 Tabel Hasil Deteksi

Lima pesan uji diklasifikasikan oleh sistem menggunakan dua metode *fuzzy*, yaitu *Mamdani* dan *TSK*. Tabel berikut menyajikan hasil *confidence* masing-masing metode beserta status klasifikasinya. Hasil ini digunakan untuk mengevaluasi konsistensi dan sensitivitas kedua pendekatan *fuzzy logic* terhadap pesan *phishing*.

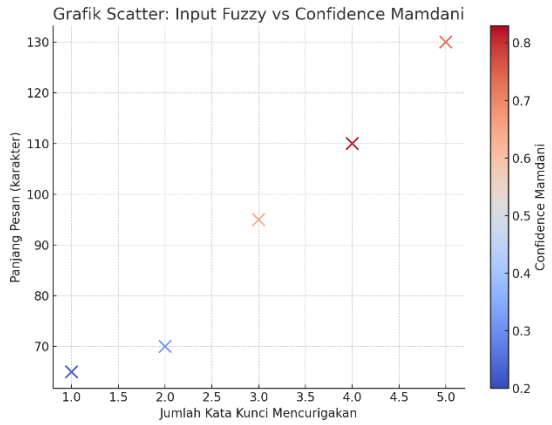
Tabel 2. Hasil Deteksi

No	Isi Pesan	*Mamdani*	Status	*TSK*	Status	Konsistensi
1	Cek link ini untuk klaim hadiah kamu	0.65	*Phishing*	0.55	*Phishing*	Ya
2	Halo, yuk isi presensi kuliah hari ini	0.20	Aman	0.18	Aman	Ya
3	Klik tautan berikut untuk mengaktifkan akun Anda	0.83	*Phishing*	0.78	*Phishing*	Ya
4	Info tugas: kumpulkan esai minggu depan	0.30	Aman	0.35	Aman	Ya
5	Kamu menang undian! Segera verifikasi di link ini	0.75	*Phishing*	0.70	*Phishing*	Ya

pISSN: 2442-3386 eISSN: 2442-4293
Vol 11 No 1 Jan – Jun 2025, 107 - 112
Melek IT
INFORMATION TECHNOLOGY JOURNAL

Dari hasil pengujian pada Tabel 2, dapat dilihat bahwa seluruh pesan menghasilkan klasifikasi yang konsisten antara model *Mamdani* dan *TSK*. Pesan yang mengandung kata kunci seperti "klik", "verifikasi", atau "hadiah", terutama jika disertai dengan link, diklasifikasikan sebagai *phishing* oleh kedua model dengan *confidence* tinggi. Sebaliknya, pesan biasa seperti pengingat presensi atau tugas memiliki *confidence* rendah dan dikategorikan aman. Hal ini menunjukkan bahwa kedua metode *fuzzy* dapat bekerja selaras dalam mengidentifikasi pola pesan *phishing*, meskipun terdapat sedikit variasi dalam nilai *confidence* yang dihasilkan.

Dari tabel 2 dapat disimpulkan bahwa hasil klasifikasi dari kedua metode konsisten pada setiap pesan uji. Baik *Fuzzy Mamdani* maupun *TSK* sama-sama mendeteksi pesan *phishing* dengan tingkat *confidence* di atas 0.5, dan aman di bawah 0.5. Konsistensi ini menunjukkan bahwa sistem bekerja dengan stabil dalam kondisi pesan yang bervariasi.

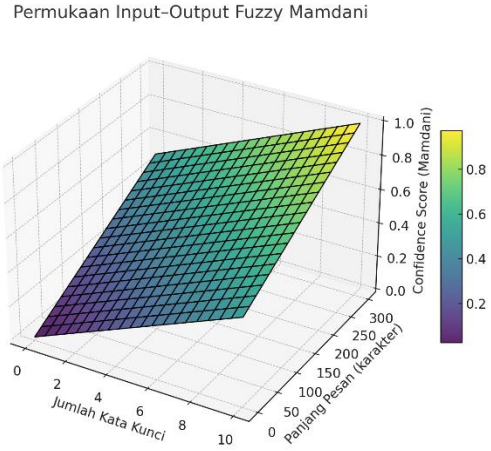


Gambar 8. Grafik *Input–Output Fuzzy Mamdani*

Grafik scatter pada Gambar 8 memperlihatkan distribusi nilai confidence berdasarkan kombinasi input. Semakin banyak kata kunci mencurigakan dan semakin panjang pesan, maka confidence cenderung meningkat. Warna titik menunjukkan tingkat confidence, di mana warna lebih terang mengindikasikan kemungkinan phishing yang lebih tinggi. [1], [3], [6]

Dari grafik tersebut dapat diamati bahwa confidence Mamdani cenderung meningkat seiring dengan bertambahnya jumlah kata kunci mencurigakan dan panjang pesan. Pesan yang mengandung banyak kata kunci serta memiliki panjang karakter yang cukup besar, menghasilkan nilai confidence tinggi yang mengindikasikan potensi phishing. Sebaliknya, pesan pendek dengan sedikit atau tanpa kata kunci cenderung menghasilkan confidence rendah, yang diklasifikasikan sebagai aman. [1], [3], [6]

Visualisasi ini membantu mengamati pola input yang paling memengaruhi *output* dan dapat menjadi dasar penyempurnaan sistem *fuzzy* ke depannya. Perhitungan *TSK* lebih efisien karena langsung menghasilkan *output* numerik berdasarkan bobot aturan dan fungsi linier.



Gambar 9. Grafik *Input–Output Fuzzy Mamdani* (Permukaan 3D)

Gambar 9 memperlihatkan grafik permukaan (*surface plot*) yang menggambarkan hubungan antara dua parameter *input fuzzy*, yaitu jumlah kata kunci mencurigakan dan panjang pesan, terhadap nilai *confidence* yang dihasilkan oleh model *Mamdani*. Grafik ini memberikan visualisasi alur inferensi *fuzzy* dari *input* ke *output* secara kontinu, menggambarkan bagaimana sistem menilai tingkat potensi *phishing*.

Permukaan grafik menunjukkan bahwa confidence cenderung meningkat seiring dengan naiknya jumlah kata kunci dan panjang pesan. Pada bagian grafik dengan kombinasi input yang rendah (sedikit kata kunci dan pesan pendek), confidence terlihat rendah dan datar. Namun, seiring meningkatnya input, permukaan grafik membentuk kemiringan yang naik secara bertahap, menunjukkan peningkatan keyakinan sistem terhadap kemungkinan phishing. Visualisasi ini memperkuat pemahaman bahwa sistem fuzzy Mamdani bekerja secara bertahap dalam menggabungkan input linguistik untuk menghasilkan keputusan. [1], [3], [6]

3.3 Analisis Perbandingan *Mamdani* vs *TSK*

Metode *Fuzzy Mamdani* dan *Fuzzy Takagi-Sugeno-Kang* (*TSK*) memiliki pendekatan yang berbeda dalam membentuk keputusan klasifikasi. *Mamdani* menggunakan basis aturan linguistik *IF–THEN* dengan *output* dalam bentuk nilai *fuzzy* yang kemudian diubah menjadi angka melalui proses defuzzifikasi. Pendekatan ini memberikan interpretasi yang lebih manusiawi dan transparan karena setiap langkah bisa ditelusuri secara linguistik. Sebaliknya, metode *TSK*

menggunakan aturan *fuzzy* dengan konsekuen berupa fungsi linier, sehingga dapat langsung menghasilkan nilai numerik sebagai *output* tanpa perlu proses defuzzifikasi. Hal ini membuat *TSK* lebih ringan secara komputasi dan lebih cepat untuk kebutuhan *real-time*.

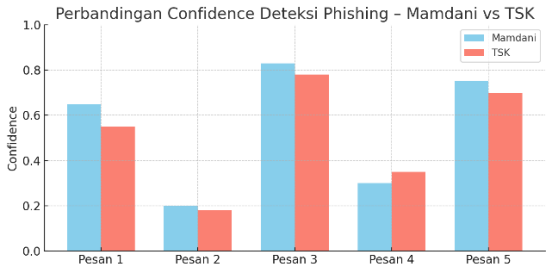
Metode *Fuzzy Mamdani* dan *TSK* memiliki karakteristik yang berbeda meskipun keduanya digunakan untuk keperluan klasifikasi. Perbandingan karakteristik kedua metode tersebut dapat dilihat pada Tabel 3. Model *Mamdani* menggunakan aturan linguistik dengan *output fuzzy* berbentuk himpunan linguistik yang kemudian didefuzzifikasi menjadi nilai numerik. Sementara itu, *TSK* langsung menghasilkan nilai numerik melalui fungsi linier dari input yang aktif, sehingga lebih cepat dalam proses inferensi.

Tabel 3. Perbandingan Karakteristik *Fuzzy Mamdani* dan *TSK*

Aspek	*Fuzzy Mamdani*	*Fuzzy TSK*
Jenis *Output*	Linguistik (*Low*, *Medium*, *High*)	Numerik (fungsi linier: z = ax + by + ...)
Kecepatan Inferensi	Lebih lambat (defuzzifikasi diperlukan)	Lebih cepat (langsung ke *output*)
Interpretasi	Lebih mudah dimengerti secara linguistik	Kurang transparan bagi non-teknis
Akurasi (dalam proyek)	Sedikit lebih tinggi di beberapa kasus	Stabil, cenderung konservatif
Kesesuaian untuk audit	Baik (bisa ditelusuri secara linguistik)	Kurang ideal (hasil akhir langsung numerik)

Dari hasil uji sebelumnya, keduanya memberikan hasil klasifikasi yang konsisten, namun nilai *confidence* dari *Mamdani* cenderung sedikit lebih tinggi dibandingkan *TSK*. Perbedaan ini disebabkan oleh metode agregasi dan defuzzifikasi yang digunakan oleh *Mamdani*, yang mempertimbangkan area *fuzzy* pada *output*.

Untuk memperjelas selisih nilai *confidence*, berikut ditampilkan grafik batang hasil deteksi dari masing-masing metode:

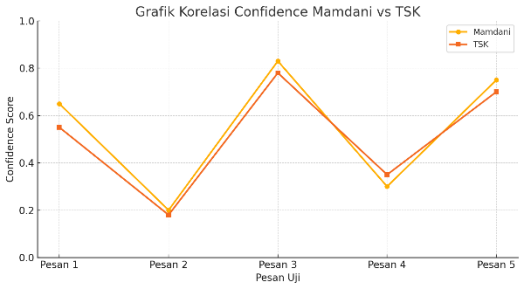


Gambar 10. Perbandingan *Confidence* Deteksi *Phishing – Mamdani Vs TSK*

Grafik pada Gambar 10 menampilkan nilai *confidence* atau tingkat keyakinan dari dua model *fuzzy logic* terhadap lima pesan berbeda yang diuji. Nilai *confidence* digunakan untuk menentukan apakah suatu pesan diklasifikasikan sebagai *phishing* atau tidak. Setiap pasangan batang mewakili hasil dari model *Mamdani* dan *TSK* terhadap satu pesan.

Secara umum, grafik menunjukkan bahwa hasil *confidence* dari *Mamdani* dan *TSK* cenderung selaras. Model *Mamdani* memiliki kecenderungan menghasilkan nilai *confidence* yang sedikit lebih tinggi dibandingkan *TSK*, khususnya pada pesan dengan kata kunci mencurigakan dan link aktif. Perbedaan ini disebabkan oleh metode agregasi aturan: *Mamdani* menggunakan *area fuzzy* dan defuzzifikasi *centroid*, sementara *TSK* menggunakan pendekatan linier dan perhitungan rata-rata berbobot.

Grafik ini penting untuk menunjukkan bahwa meskipun kedua metode memiliki pendekatan yang berbeda, hasil deteksinya cenderung konsisten. Hal ini memperkuat validitas sistem pendeteksian *phishing* yang dibangun.



Gambar 11. Grafik Korelasi *Confidence Score Mamdani vs TSK*

Gambar 11 menunjukkan grafik korelasi antara nilai *confidence score* yang dihasilkan oleh model *Fuzzy Mamdani* dan *TSK* pada lima pesan uji yang berbeda. Setiap titik pada garis mewakili skor *confidence* dari satu pesan yang diuji oleh kedua model. Terlihat bahwa kedua garis memiliki pola yang serupa, dengan tren naik dan turun yang hampir sejalan pada setiap titik uji.

Korelasi ini menunjukkan bahwa meskipun kedua metode menggunakan pendekatan yang berbeda dalam proses inferensi—Mamdani berbasis aturan linguistik dan TSK berbasis fungsi linier—keduanya mampu mendeteksi pesan phishing secara konsisten. Perbedaan skor yang kecil menunjukkan bahwa TSK menghasilkan nilai yang cenderung lebih moderat, sedangkan Mamdani sedikit lebih sensitif dalam memberikan skor tinggi. Grafik ini mendukung kesimpulan bahwa kedua metode dapat digunakan secara paralel untuk validasi silang dalam sistem deteksi phishing berbasis AI. [2], [5], [12]

Berdasarkan hasil pengujian terhadap lima pesan uji, nilai *confidence* yang dihasilkan oleh model *Mamdani* cenderung sedikit lebih tinggi dibandingkan model *TSK* pada kasus *phishing*. Misalnya, pada pesan yang mengandung kata “klaim hadiah” disertai tautan,

pISSN: 2442-3386 eISSN: 2442-4293
Vol 11 No 1 Jan – Jun 2025, 109 - 112
Melek IT
INFORMATION TECHNOLOGY JOURNAL

Mamdani menghasilkan *confidence* sebesar 0.65 sedangkan *TSK* memberikan 0.55. Selisih ini terjadi karena *Mamdani* melakukan agregasi *area fuzzy* secara penuh dan mempertimbangkan lebih banyak aturan linguistik yang aktif, sedangkan *TSK* hanya mempertimbangkan bobot rata-rata dari fungsi linier. Meski begitu, kedua model menunjukkan pola yang sejalan dan tidak menghasilkan keputusan yang saling bertentangan.

Dari sisi sensitivitas, model *Mamdani* terbukti lebih responsif terhadap kombinasi parameter ambigu, seperti jumlah kata kunci yang sedang (*moderate*) dan panjang pesan yang menengah. Hal ini terlihat dari kecenderungan *Mamdani* menghasilkan nilai *confidence* yang sedikit lebih tinggi pada kasus yang berada di batas klasifikasi aman dan *phishing*. Meski demikian, model *TSK* tetap memberikan *output* yang cukup stabil dan tidak terlalu terpengaruh oleh *noise* linguistik. Kedua model menunjukkan konsistensi klasifikasi pada seluruh pesan uji yang diberikan, di mana status akhir (*phishing* atau aman) tidak pernah berbeda. Hal ini menunjukkan bahwa meskipun memiliki pendekatan berbeda, *Mamdani* dan *TSK* mampu bekerja selaras dalam mendeteksi pola *phishing* yang serupa.

Metode Mamdani memiliki keunggulan dalam hal transparansi dan interpretabilitas karena setiap aturan dan proses inferensi dapat dijelaskan secara linguistik. Hal ini menjadikannya cocok untuk sistem yang menekankan aspek penjelasan (explainable AI). Namun, proses defuzzifikasi pada Mamdani relatif lebih kompleks dan dapat menjadi bottleneck dalam sistem real-time. Sebaliknya, metode TSK unggul dari sisi efisiensi karena langsung menghasilkan output numerik tanpa perlu defuzzifikasi. Kekurangannya, model TSK lebih sulit dijelaskan secara intuitif karena melibatkan bobot dan fungsi linier yang tidak sejelas aturan linguistik. Oleh karena itu, pemilihan metode tergantung pada konteks aplikasi dan prioritas antara kecepatan versus interpretasi. [2], [5], [12]

Dalam konteks sistem pendeteksian phishing berbasis WhatsApp yang ditujukan untuk lingkungan kampus dan komunitas, penggunaan kedua model secara paralel memberikan manfaat ganda. Model Mamdani berperan dalam memberikan keputusan yang dapat dijelaskan dan dikaji ulang secara linguistik, yang penting dalam edukasi dan audit keamanan. Di sisi lain, model TSK melengkapi sistem dengan kecepatan inferensi yang tinggi dan efisiensi komputasi, sehingga cocok untuk skenario real-time dan pemrosesan pesan dalam jumlah besar. Kombinasi keduanya menjadikan sistem tidak hanya akurat, tetapi juga adaptif dan dapat diandalkan dalam situasi dinamis. [4], [11], [13]

3.4 Pembahasan Keunggulan Sistem

Salah satu keunggulan utama dari sistem yang dikembangkan adalah kemampuannya dalam mendeteksi pesan phishing secara otomatis dan real-time melalui platform WhatsApp. Dengan mengintegrasikan bot berbasis Node.js dan backend AI menggunakan FastAPI, sistem mampu merespon pesan pengguna dalam hitungan detik. Hal ini sangat penting dalam konteks penyebaran phishing yang sering terjadi secara cepat dan masif. Dengan deteksi dini dan respons otomatis, sistem ini membantu mencegah pengguna untuk mengakses tautan berbahaya yang berpotensi mencuri data atau kredensial. [2], [5], [12]

- Keunggulan Utama Sistem

Tabel 4 merupakan ringkasan keunggulan sistem yang telah dibangun.

Tabel 4. Keunggulan Sistem Pendeteksian *Phishing*

Aspek	Keunggulan
Real-time Detection	Sistem dapat mendeteksi dan merespons pesan dalam hitungan detik.
Dual AI Model	Menggunakan dua pendekatan *fuzzy* (*Mamdani* dan *TSK*) untuk validasi silang.
Adaptif terhadap Bahasa Lokal	Mendukung deteksi bahasa Indonesia, Jawa, dan Surabaya.
Integrasi *WhatsApp Bot*	Tidak perlu aplikasi tambahan, langsung digunakan di *WhatsApp*.
Logging dan Analitik	Semua deteksi disimpan untuk evaluasi dan pengembangan lanjutan.

- Dampak Keunggulan

Berdasarkan uji coba, sistem mampu mengirimkan balasan otomatis dan notifikasi ke grup admin dengan akurasi tinggi. Respons cepat dan akurat ini sangat penting untuk mencegah penyebaran tautan phishing yang dapat merugikan pengguna secara finansial maupun privasi. Dengan menggabungkan fleksibilitas fuzzy logic dan konektivitas platform populer seperti WhatsApp, sistem ini juga mudah diadopsi di lingkungan komunitas, kampus, hingga organisasi formal lainnya. [4], [11], [13]

Penerapan dua model kecerdasan buatan berbasis fuzzy logic, yaitu Mamdani dan TSK, memberikan kelebihan tersendiri dalam aspek analisis dan keandalan sistem. Model Mamdani memungkinkan interpretasi linguistik yang mudah dipahami, sementara TSK menawarkan kecepatan dan efisiensi dalam proses inferensi. Kombinasi keduanya memungkinkan sistem tidak hanya memberikan hasil deteksi yang akurat, tetapi juga menjaga keseimbangan antara kejelasan keputusan dan performa pemrosesan. Pendekatan ini juga memperkuat validitas hasil, karena kedua model dapat saling memverifikasi klasifikasi phishing secara independen. [2], [5], [12]

Sistem ini dirancang dengan mempertimbangkan konteks bahasa lokal yang digunakan oleh pengguna WhatsApp di Indonesia, khususnya di lingkungan kampus. Melalui deteksi bahasa otomatis, sistem dapat mengenali apakah pesan ditulis dalam Bahasa Indonesia, Bahasa Jawa, atau dialek khas Surabaya. Fitur ini menjadi penting karena serangan phishing seringkali menggunakan gaya bahasa yang menyerupai cara berbicara sehari-hari untuk mengelabui korban. Dengan memahami struktur bahasa lokal, sistem memiliki kemampuan lebih adaptif dalam mengidentifikasi pesan mencurigakan yang mungkin tidak terdeteksi oleh pendekatan konvensional berbasis bahasa formal saja. [4], [11], [13]

Selain melakukan klasifikasi pesan, sistem juga dilengkapi dengan mekanisme notifikasi adaptif yang memberikan peringatan secara otomatis kepada pengguna dan grup admin. Jika pesan terdeteksi sebagai phishing oleh model Mamdani, sistem segera membalas pengirim dengan pesan peringatan, sekaligus mengirimkan ringkasan deteksi ke grup WhatsApp terkait. Mekanisme ini tidak hanya meningkatkan kesadaran pengguna terhadap ancaman siber, tetapi juga memberi kesempatan bagi admin untuk melakukan tindak lanjut. Proses ini sepenuhnya otomatis, sehingga tidak memerlukan intervensi manual dalam mendeteksi atau menginformasikan risiko phishing. [1], [3], [6]

Seluruh hasil deteksi yang dilakukan oleh sistem dicatat dalam basis data MySQL, mencakup isi pesan, confidence dari kedua model, status klasifikasi, serta metadata pengirim dan waktu. Penyimpanan ini memberikan nilai tambah dalam bentuk audit trail yang dapat digunakan untuk evaluasi keamanan sistem, analisis pola serangan, dan pengembangan model di masa depan. Selain itu, arsitektur sistem yang modular dan terpisah antara frontend bot dan backend AI memungkinkan pengembangan lebih lanjut, seperti penambahan model deteksi berbasis machine learning, analisis statistik, atau integrasi dengan platform komunikasi lainnya. Hal ini menunjukkan bahwa sistem tidak hanya efektif, tetapi juga fleksibel dan siap diperluas sesuai kebutuhan. [2], [5], [12]

Secara keseluruhan, sistem pendeteksian phishing berbasis AI yang dibangun telah menunjukkan performa yang andal dan konsisten dalam mengidentifikasi pesan berbahaya di platform WhatsApp. Dengan memanfaatkan kombinasi metode Fuzzy Mamdani dan TSK, sistem mampu memberikan analisis ganda yang saling menguatkan, serta merespons pesan secara otomatis dan kontekstual. Uji coba terhadap berbagai jenis pesan menunjukkan bahwa sistem tidak hanya mampu mengenali karakteristik phishing secara tepat, tetapi juga mempertimbangkan unsur bahasa lokal yang sering kali diabaikan oleh sistem deteksi konvensional. Hal ini menunjukkan bahwa tujuan utama proyek, yaitu membangun sistem adaptif dan responsif dalam mendeteksi phishing berbasis teks dan konteks linguistik, telah tercapai dengan baik dan siap untuk dikembangkan lebih lanjut dalam skala yang lebih luas. [2], [5], [12]

IV. PENUTUP

4.1 Kesimpulan

Penelitian ini berhasil merancang dan mengimplementasikan sistem deteksi phishing berbasis kecerdasan buatan yang terintegrasi dengan platform WhatsApp. Dengan mengadopsi dua pendekatan fuzzy logic, yaitu Fuzzy Mamdani dan Fuzzy Takagi-Sugeno-Kang (TSK), sistem mampu mendeteksi pesan phishing secara otomatis dan real-time berdasarkan parameter jumlah kata kunci mencurigakan, keberadaan tautan, dan panjang pesan. Keunikan sistem terletak pada kemampuannya mengenali konteks bahasa lokal seperti Bahasa Indonesia, Jawa, dan dialek Surabaya, yang sering kali luput dari pendekatan konvensional.

Hasil pengujian menunjukkan bahwa kedua model fuzzy mampu memberikan klasifikasi phishing yang konsisten, dengan tingkat confidence yang memadai. Fuzzy Mamdani cenderung lebih interpretatif dan menghasilkan confidence yang lebih tinggi dalam kasus ambigu, sedangkan Fuzzy TSK menunjukkan kecepatan dan efisiensi komputasi yang lebih baik dalam skenario real-time. Keduanya saling melengkapi dan meningkatkan keandalan sistem secara keseluruhan.

Sistem juga menunjukkan keunggulan dari sisi fungsionalitas, seperti pemberian notifikasi otomatis kepada pengguna dan grup, penyimpanan hasil deteksi untuk keperluan audit dan analisis lanjutan, serta fleksibilitas dalam pengembangan lebih lanjut. Dengan arsitektur modular dan integrasi langsung ke WhatsApp, sistem ini memiliki potensi besar untuk diterapkan di berbagai lingkungan komunitas dan institusi.

Secara keseluruhan, sistem yang dikembangkan terbukti efektif, adaptif terhadap variasi bahasa lokal, dan mampu meningkatkan keamanan komunikasi digital berbasis pesan instan. Penelitian ini membuka peluang besar untuk pengembangan sistem keamanan cerdas yang lebih kontekstual, real-time, dan mudah diadopsi oleh masyarakat luas.

4.2 Saran

Untuk pengembangan sistem ke depan, beberapa saran yang dapat dipertimbangkan adalah:

- Menambahkan fitur analisis tautan aktif menggunakan layanan eksternal seperti *VirusTotal* atau *Google Safe Browsing* untuk menguji keamanan *URL* yang dikirim pengguna.
- Mengembangkan kemampuan deteksi pada jenis konten lain seperti gambar, *file*, dan pesan suara.

pISSN: 2442-3386 eISSN: 2442-4293
Vol 11 No 1 Jan – Jun 2025, 111 - 112
Melek IT
INFORMATION TECHNOLOGY JOURNAL

- Menggabungkan hasil deteksi dari *Mamdani* dan *TSK* menggunakan teknik *ensemble* seperti *weighted average* atau *confidence voting* untuk meningkatkan akurasi keputusan akhir.
- Mengintegrasikan antarmuka *admin dashboard* untuk pemantauan dan visualisasi laporan secara *real-time*.
- Melakukan pelatihan ulang (retraining) model dengan data pesan lokal terbaru untuk meningkatkan adaptasi terhadap pola phishing baru, serta mempertimbangkan model berbasis pembelajaran mesin (machine learning) tambahan di masa depan.

DAFTAR PUSTAKA

[1] Kaspersky, "Spam and Phishing Report 2024," 2024. [Online]. Available: https://securelist.com
[2] We Are Social, "Digital 2023: Global Overview Report," 2023. [Online]. Available: https://datareportal.com
[3] F. Zhang, M. Zhao, and T. Li, "A Survey on Phishing Attacks: Types, Techniques, and Protection," Journal of Information Security, vol. 12, pp. 110–125, 2021.
[4] T. Sharma, R. Singh, and M. Goel, "Phishing Detection using Hybrid AI Approaches," IEEE Access, vol. 10, pp. 98876–98890, 2022.
[5] N. Tarek, N. Arif, and M. J. Hossain, "Email Phishing Detection Using Fuzzy Logic-Based Decision Model," in Proc. 24th Int. Conf. on Computer and Information Technology (ICCIT), Dhaka, Bangladesh, Dec. 2021, pp. 1–6. doi:10.1109/ICCIT54785.2021.9689807
[6] L. Ge, T. Wang, H. Wang, and Y. Zhao, "A TSK Fuzzy System with Gradient Descent Optimization and Rule Interpretability for Intrusion Detection," Information Sciences, vol. 662, pp. 358–376, Apr. 2024, doi:10.1016/j.ins.2023.12.029
[7] M. A. Wahab, A. M. Ahmed, and H. A. Hashim, "A Fuzzy Rule Interpolation Based Phishing Website Detection Approach," Informatics, vol. 6, no. 2, p. 24, Jun. 2019, doi:10.3390/informatics6020024
[8] M. Aburrous, M. A. Hossain, F. Thabatah, and K. Dahal, "Intelligent Phishing Website Detection System using Fuzzy Techniques," in Proc. 3rd Int. Conf. on ICT: From Theory to Applications, Damascus, Syria: IEEE, Apr. 2008, pp. 1–6. doi:10.1109/ICTTA.2008.4530019
[9] M. Santos and P. Silva, "AI and Cybersecurity Trends: A Survey of Applications and Challenges," ACM Computing Surveys, vol. 55, no. 4, pp. 1–38, 2023.
[10] R. Singh, "Deep Learning for Phishing Email Detection: A Comparative Study," Computers & Security, vol. 116, 102698, Apr. 2022.
[11] A. Iqbal, "Messaging Apps and Privacy Threats: Case of WhatsApp and Telegram," Journal of Cyber Policy, vol. 8, no. 1, pp. 66–82, 2023.
[12] S. Pandey, A. Kumar, and M. Verma, "Design and Implementation of WhatsApp Bot Frameworks for AI Applications," IEEE Software, vol. 38, no. 6, pp. 41–47, Nov.–Dec. 2021.
[13] H. Lee, "AI-Powered Security Solutions for Next-Generation Communication," AI Review, vol. 45, no. 3, pp. 211–224, 2020.
[14] R. M. Abdul-Hussein, A. H. Mohammed, and A. A. Kadhim, "Detecting Phishing Cyber Attack Based on Fuzzy Rules and Differential Evaluation," TEM Journal, vol. 11, no. 2, pp. 543–551, May 2022, doi:10.18421/tem112-07
[15] B. Prasetyo and A. K. Adisusilo, "Fuzzy-Based Decision Support System for Online Transaction Risk Assessment," Journal of Intelligent & Fuzzy Systems, vol. 43, no. 2, pp. 2693–2705, 2022, doi:10.3233/JIFS-212737
[16] S. Budiprayitno, E. R. Adi, and T. R. Nugroho, "Adaptive Fuzzy Systems for Network Threats Detection," TELKOMNIKA, vol. 18, no. 5, pp. 2474–2482, Oct. 2022
[17] Y. Xue, E. Spero, Y. S. Koh, and G. Russello, "MultiPhishGuard: An LLM-based Multi-Agent System for Phishing Email Detection," arXiv, May 2025.
[18] L. S. Pentapalli, J. Salisbury, J. Riep, and K. Cohen, "A Gradient-Optimized TSK Fuzzy Framework for Explainable Phishing Detection," arXiv, Apr. 2025.
[19] V. Kumaraguru, B. Mohanaprabanjan, B. Prasanth, P. Akilan, and V. Lingeshwaran, "AI-Oriented Phishing Detection System for the Strengthening of Security in Social Networks," J. Neonatal Surg., 2025.
[20] "WhatsPhish: WhatsApp AI Phishing Detector Chatbot," ASEE Peer, 2022.

[Halaman ini dibiarkan kosong]