

Prediksi Retweet Berdasarkan Fitur User-Based, Content-Based, dan Time-Based Menggunakan Metode ANN-GSO

1st Raisul Muhalani
Fakultas Informatika
Universitas Telkom
Bandung, Indonesia
raisulmuha@students.telkomuniversity.
ac.id

2nd Jondri
Fakultas Informatika
Universitas Telkom
Bandung, Indonesia
jondri@telkomuniversity.ac.id

3rd Indwiarti
Fakultas Informatika
Universitas Telkom
Bandung, Indonesia
indwiarti@telkomuniversity.ac.id

Abstrak - Media sosial menjadi salah satu platform yang banyak dipilih untuk sarana saling berbagi informasi, hiburan, serta dapat membuat mereka menghilangkan rasa penat dari aktifitas mereka sehari-hari. Media sosial sudah menjadi kebutuhan untuk sebagian besar masyarakat khususnya Indonesia. Salah satu media sosial yang sering digunakan oleh masyarakat Indonesia yaitu Twitter. Twitter dapat membagikan sebuah postingan yang biasa disebut dengan tweet (kicauan) yang dapat digunakan oleh pengguna untuk membagikan tulisan, foto, video, maupun gif kepada publik. Salah satu fitur Twitter yaitu retweet. Fitur retweet ini memiliki fungsi untuk membagikan kembali sebuah postingan, baik postingan mereka sendiri maupun postingan pengguna lain. Fitur ini sangat berperan penting dalam penyebaran informasi. Penelitian ini membahas mengenai prediksi retweet menggunakan fitur user-based, content-based, dan time-based dengan metode Jaringan Saraf Tiruan (Artificial Neural Network) untuk klasifikasinya, yang dioptimalkan dengan algoritma Glowworm Swarm Optimization (GSO) untuk mendapatkan tingkat akurasi yang lebih tinggi. Model ANN yang dioptimalkan dengan GSO menunjukkan hasil terbaik ketika dilakukan skenario oversampling, dengan akurasi sebesar 78% dan F1-Score 78%. Pada GSO terdapat peningkatan pada dataset model prediksi secara keseluruhan.

Kata kunci - klasifikasi, Twitter, retweet, ANN, GSO

I. PENDAHULUAN

A. Latar Belakang

Media sosial telah menjadi bagian penting dari kehidupan sehari-hari banyak orang yang terus berkembang seiring dengan penambahan usia pengguna. Platform seperti Facebook, Twitter, Instagram, dan lainnya menyediakan berbagai cara bagi pengguna untuk berinteraksi, berbagi informasi, dan mengikuti perkembangan terbaru [1]. Seiring waktu, media sosial tidak hanya berfungsi sebagai alat komunikasi, tetapi juga sebagai sarana pemasaran, analisis bisnis, dan berbagai keperluan lainnya [2].

Media sosial telah menjadi bagian dari kehidupan orang, bersama dengan perkembangan usia, media sosial terus berevolusi [3]. Keberadaannya membuat banyak informasi bisa mudah untuk mendapatkan untuk

pengguna. Salah satu media sosial yang cukup banyak penggunaannya adalah Twitter.

Pada Twitter, setiap pengguna dapat membuat profil, menulis pesan, dan berbagi informasi dengan pengguna lain. Melalui tweet, informasi dapat disebarkan ke publik secara real-time, dengan tweet yang dapat berisi foto, video, atau tautan. Twitter juga memiliki fitur retweet yang memungkinkan pengguna untuk memposting ulang tweet orang lain atau tweet mereka sendiri untuk dibagikan kepada followers. Fitur Retweet ini menjadi kunci dalam mekanisme difusi informasi, menjelaskan bagaimana informasi dapat menyebar secara luas. Indonesia merupakan negara peringkat ketiga sebagai pengguna Twitter.

Twitter memiliki format baru berbeda dengan Facebook yaitu berupa microblogging, memiliki 280 karakter tulisan untuk setiap tweet atau cuitannya, pada awalnya hanya 140 karakter tulisan di setiap tweetnya namun dianggap terlalu sedikit. Hal tersebut memungkinkan membantu penggunaannya untuk berbagi informasi [4].

Oleh karena itu pada penulisan penelitian dalam bentuk tugas akhir ini, peneliti membangun sebuah model Prediksi retweet berdasarkan fitur user-based, content-based, dan time-based menggunakan metode ANN-GSO.

B. Topik dan Batasannya

Berdasarkan latar belakang, berikut ini adalah topik masalahnya yaitu prediksi retweet berdasarkan fitur user-based, content-based, dan time-based menggunakan metode ANN-GSO. Sedangkan untuk batasannya, dataset yang digunakan yaitu U-20 sejumlah 869 data. Dataset U-20 merupakan data yang diambil dari Twitter dengan kata kunci piala dunia usia 20.

C. Tujuan

Tujuan dari penelitian ini yaitu membuat model dan mengukur tingkat akurasi prediksi retweet berdasarkan fitur user-based, content-based, dan time-based menggunakan metode ANN-GSO pada Twitter.

D. Organisasi Tulisan

Bagian selanjutnya akan menjelaskan studi terkait,

berdasarkan penelitian dan hasilnya. Kemudian, pada bagian berikutnya, akan dijelaskan urutan sistem yang dibangun dalam penelitian ini. Selanjutnya, bagian evaluasi akan membahas hasil penelitian. Terakhir, bagian kesimpulan akan menyajikan ringkasan penelitian beserta saran untuk penelitian selanjutnya.

II. STUDI TERKAIT

Pada penelitian ini, peneliti menggunakan beberapa penelitian relevan yang terkait dengan fokus penelitian. Penelitian relevan ini menjadi acuan bagi peneliti, dalam menjalankan penelitian pengembangan.

Penelitian relevan pertama oleh Edvan et al. terkait kemajuan teknologi telah menyebabkan penyebaran informasi yang cepat, terutama di media sosial seperti Twitter. Retweeting, atau posting ulang pesan, dianggap sebagai mekanisme penyebaran informasi yang mudah diakses yang disediakan oleh Twitter. Dengan memahami alasan seorang pengguna me-retweet tweet orang lain dan dengan membuat prediksi ini, kita dapat memahami bagaimana informasi menyebar di Twitter. Dalam penelitian ini, Artificial Neural Network – Genetic Algorithm digunakan untuk proses klasifikasi dengan fitur berbasis pengguna dan konten. Hasil evaluasi yang diperoleh adalah 90% akurasi, 72% presisi, 83% recall, dan 65% nilai F1-Score pada model dengan oversampling[5].

Penelitian relevan kedua oleh Yahya et al. terkait transportasi online membahas inovasi dalam transportasi yang muncul bersamaan dengan pengembangan aplikasi online yang menyediakan berbagai fitur dan kenyamanan. Banyak pengguna menulis tanggapan mereka di media sosial seperti Twitter. Pendapat dan tanggapan ini seringkali ditransmisikan langsung oleh pengguna transportasi online ke akun resmi mereka. Tanggapan yang diberikan oleh pengguna sangat banyak dan dapat digunakan untuk analisis sentimen terhadap transportasi online. Namun, proses analisis ini tidak dapat dilakukan secara manual. Oleh karena itu, diperlukan sistem yang dapat membantu menganalisis tanggapan pengguna di Twitter secara otomatis. Dalam penelitian ini, sebuah sistem analisis sentimen dibangun untuk transportasi online di Indonesia menggunakan algoritma ensemble stacking, yang menyederhanakan dan meningkatkan akurasi analisis sentimen. Ensemble stacking merupakan solusi untuk metode pembelajaran mesin canggih yang dapat meningkatkan kinerja klasifikator dasar. Sistem yang dibangun menggunakan ensemble stacking dengan tiga klasifikator dasar, yaitu SVM kernel RBF, SVM kernel linear, dan regresi logistik. Hasil akurasi terbaik pada dataset Gojek adalah 88%, dan skor F1 terbaik adalah 87%. Ensemble stacking yang diterapkan dalam penelitian ini pada analisis sentimen transportasi online di Twitter menghasilkan akurasi yang lebih baik dibandingkan dengan klasifikator dasar yang digunakan[6].

Penelitian ketiga oleh Artamira dilakukan dengan mengumpulkan data menggunakan API Twitter. Analisis sentimen digunakan untuk menentukan distribusi

perasaan positif dan negatif. Setelah distribusi tweet diketahui, klasifikasi dilakukan menggunakan metode ensemble bagging untuk menilai kinerja metode tersebut. Metode klasifikasi menggunakan ensemble bagging dengan tiga algoritma dasar, yaitu Naive Bayes, K-Nearest Neighbor, dan Decision Tree. Sementara itu, ekstraksi fitur yang digunakan dalam penelitian ini adalah TF-IDF (Term Frequency-Inverse Document Frequency). Kinerja metode ensemble bagging dengan menerapkan hyperparameter tuning mencapai akurasi 0,72, recall 0,71, F1-Score 0,72, dan precision 0,72[7].

Penelitian keempat oleh Jondri et al. membahas tentang retweet, yaitu cara untuk menyebarkan informasi di Twitter. Sebuah tweet dipengaruhi oleh beberapa fitur yang menentukan apakah tweet tersebut akan di-retweet atau tidak. Penelitian ini membahas fitur-fitur yang memengaruhi penyebaran tweet, meliputi fitur berbasis pengguna, berbasis waktu, dan berbasis konten. Fitur berbasis pengguna terkait dengan siapa yang membuat tweet, fitur berbasis waktu terkait dengan kapan tweet diunggah, sedangkan fitur berbasis konten terkait dengan isi tweet. Klasifikator yang digunakan untuk memprediksi apakah tweet akan di-retweet adalah Multi-Layer Perceptron (MLP) dan MLP yang dioptimalkan oleh algoritma kecerdasan swarm. Dalam penelitian ini, data dari pengguna Twitter Indonesia dengan hashtag FIFA U-20 digunakan. Hasil penelitian menunjukkan bahwa fitur yang paling berpengaruh dalam menentukan apakah tweet akan di-retweet atau tidak adalah fitur berbasis konten. Selain itu, ditemukan bahwa MLP yang dioptimalkan dengan algoritma kecerdasan swarm memiliki kinerja yang lebih baik dibandingkan dengan MLP biasa[8].

Penelitian relevan terakhir oleh Daffa membahas karakter publik yang sering diperhatikan oleh pengguna media sosial, baik karena apa yang mereka katakan atau karena peran mereka dalam serial televisi. Umumnya, tokoh publik mengunggah sesuatu ke akun media sosial mereka untuk membantu membentuk citra mereka. Namun, tidak semua orang yang melihat unggahan tersebut merasa senang. Beberapa bahkan tidak menyukainya. Penelitian ini bertujuan untuk menentukan perasaan masyarakat terhadap tokoh publik Anya Geraldine yang disampaikan di Twitter dalam bahasa Indonesia. Proses klasifikasi dalam penelitian ini menggunakan metode Adaptive Boosting (AdaBoost) dan Extreme Gradient Boosting (XGBoost) dengan preprocessing teks yang mencakup pembersihan, normalisasi kasus, tokenisasi, dan filtrasi. Data yang digunakan adalah tweet dalam bahasa Indonesia dengan kata kunci “@anyaselalubenaar”, dengan total 7.475 tweet, terdiri dari 6.887 tweet positif dan 588 tweet negatif. Proses oversampling digunakan untuk menghindari masalah overfitting. Fitur yang digunakan dalam penelitian ini adalah TF-IDF. Empat skenario eksperimental dilakukan untuk memvalidasi efektivitas model: kinerja model pertama tanpa oversampling, kinerja model kedua dengan oversampling, kinerja model ketiga dengan undersampling, dan kinerja model keempat dengan tuning hyperparameter. Hasil

eksperimen menunjukkan bahwa kombinasi XGBoost, SMOTE, dan hyperparameter tuning mencapai akurasi 95% dibandingkan dengan kombinasi AdaBoost, SMOTE, dan hyperparameter tuning yang mencapai akurasi 87%. Penerapan SMOTE dan hyperparameter tuning terbukti efektif dalam mengatasi masalah ketidakseimbangan data dan meningkatkan hasil klasifikasi[9], [10].

A. Twitter

Penelitian dan industri telah memperhatikan penyebaran informasi di Twitter. Meskipun tweet di Twitter terbatas pada 280 karakter, ada berbagai elemen informasi yang dapat digunakan sebagai prediktor, seperti penggunaan simbol @ untuk menandai pengguna lain, tagar (#), serta media seperti foto atau video.[10].

Twitter adalah situs jejaring sosial yang memungkinkan penggunaannya bertukar teks dengan batas karakter. Kehadiran Twitter memungkinkan pengguna menyuarakan pendapat mereka dengan bebas. Selain digunakan untuk menyuarakan opini, Twitter juga dapat digunakan untuk membaca berita. Pengguna yang ingin mendapatkan pembaruan atau berita dari pengguna lain harus mengikuti akun tersebut [11]

Menurut Putra, Twitter memiliki beberapa kelebihan dibandingkan media sosial lainnya. Jangkauannya sangat luas, tidak hanya mencakup teman, tetapi juga dapat menjangkau publik figur. Potensi periklanan di masa mendatang juga lebih besar. Komunikasi di Twitter berlangsung sangat cepat (real-time), bersifat multilink (terhubung dengan banyak jaringan), dan lebih terukur dibandingkan media sosial lainnya. Twitter membantu penyebaran informasi dengan cepat, yang kemudian menjadi topik pembahasan oleh penggunaannya [12].

Menurut Badjar et al. Twitter tidak hanya digunakan untuk mengungkapkan ide dan emosi melalui tweet, tetapi juga untuk berbagi informasi faktual dan pengetahuan yang berharga. Dalam konteks penyebaran informasi yang cepat di Twitter, sangat bermanfaat jika semua pengguna berkontribusi dengan membagikan tweet yang memiliki nilai tambah bagi seluruh pengguna. Twitter memuat beragam jenis tweet, termasuk pembaruan berita, pesan inspiratif, pandangan tentang berbagai topik, konten komedi, penafsiran ayat suci, dan kontribusi beragam lainnya yang dibuat oleh para pengguna.. [13].

Penggunaan Twitter sebagai teknologi komunikasi telah menunjukkan bahwa Twitter memperkuat teknologi komunikasi sebelumnya, yaitu komunikasi dua arah antara media massa dan audiens yang terhubung melalui internet. Twitter memungkinkan interaksi langsung dan real-time, yang membuat komunikasi lebih dinamis dan responsif dibandingkan dengan teknologi komunikasi tradisional [14].

B. ANN

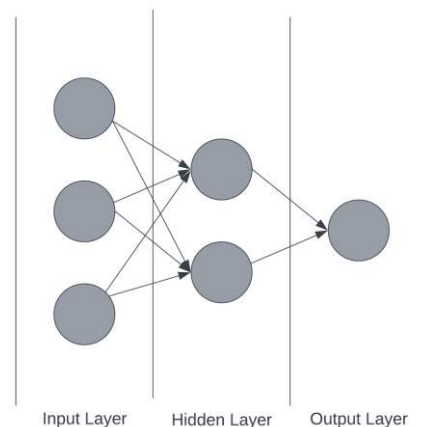
Menurut Garrett, Artificial Neural Network (ANN) adalah metode komputasi yang dirancang untuk meniru cara otak manusia mentransfer informasi. Secara umum, ANN adalah teknik yang digunakan untuk mendekati

fungsi dalam situasi di mana hubungan antara input dan output sangat rumit dan tidak linier. ANN dirancang sebagai model jaringan yang tersebar, meniru aktivitas dan fungsi kompleks otak manusia [15].

(Kustono dan Hatmojo mengakan Artificial Neural Network (ANN) atau Jaringan Syaraf Tiruan adalah sistem tiruan yang didasarkan pada struktur saraf otak manusia. Otak memiliki kemampuan untuk belajar dari pengalaman. Meskipun cara kerja otak belum sepenuhnya dipahami, perannya sebagai prosesor yang luar biasa sudah diketahui. Unsur utama otak adalah sel-sel saraf, yang juga dimiliki oleh bagian tubuh lainnya. Sel-sel saraf otak mampu mengingat, berpikir, dan menerapkan pengalaman yang telah diperoleh [16].

Jaringan Saraf Tiruan (JST) atau Artificial Neural Network adalah pendekatan yang berbeda dari metode kecerdasan buatan (AI) lainnya. JST adalah model kecerdasan yang terinspirasi dari struktur otak manusia dan diimplementasikan menggunakan program komputer yang mampu menyelesaikan berbagai proses perhitungan selama proses pembelajaran berlangsung [17]. Dapat dilihat pada Gambar 1 arsitektur dari Artificial Neural Network (ANN) terdiri dari tiga lapisan utama, yaitu:

1. Input Layer: Lapisan ini berfungsi untuk membawa data masuk ke dalam sistem untuk kemudian diproses pada lapisan berikutnya.
2. Hidden Layer: Lapisan ini terletak di antara input layer dan output layer. Pada lapisan ini, neuron-neuron buatan menerima sekumpulan input yang telah diberikan bobot (weight) dan menjalankan prosedur untuk menghasilkan output melalui fungsi aktivasi (activation function).
3. Output Layer: Lapisan terakhir dari ANN yang menghasilkan output sistem berdasarkan proses yang dilakukan pada lapisan sebelumnya.



GAMBAR 1.
Arsitektur ANN

C. Confusion matrix

Confusion matrix merupakan evaluasi yang paling sering digunakan untuk sebuah klasifikasi. Ini disebabkan akurasi dapat digunakan untuk mengevaluasi

sebuah sistem yang sedang dibangun. Untuk dapat menggambarkan kerja suatu sistem dapat menggunakan confusion matrix. Confusion matrix adalah alat analitik yang membantu kita untuk membandingkan hasil prediksi model dengan nilai sebenarnya. Alat ini bisa digunakan untuk mengukur kinerja model melalui berbagai metrik evaluasi, seperti akurasi, presisi, recall, dan F1-Score (atau F-Measure). Dalam confusion matrix, terdapat empat hasil utama yang dihasilkan, yaitu True Positive (TP), False Positive (FP), False Negative (FN), dan True Negative (TN)

TABEL 1.
Confusion matrix

		Nilai Aktual	
		Positif	Negatif
Nilai Prediksi	Positif	True Positive (TP)	False Positive (FP)
	Negatif	False Negative (FN)	True Negative (TN)

1. Precision

Precision mengidentifikasi frekuensi jawaban yang benar. Rumus precision seperti persamaan (1)

$$Precision = \frac{tp}{tp+fp} \quad (1)$$

Dimana tp adalah true positive dan fp adalah false positive. Sederhanya precision bertujuan untuk memahami bagian mana dari semua prediksi positif yang paling benar.

2. Recall

Recall mengidentifikasi frekuensi pendeteksian. Rumus recall seperti persamaan (2)

$$Recall = \frac{tp}{tp+fn} \quad (2)$$

3. F1-Score

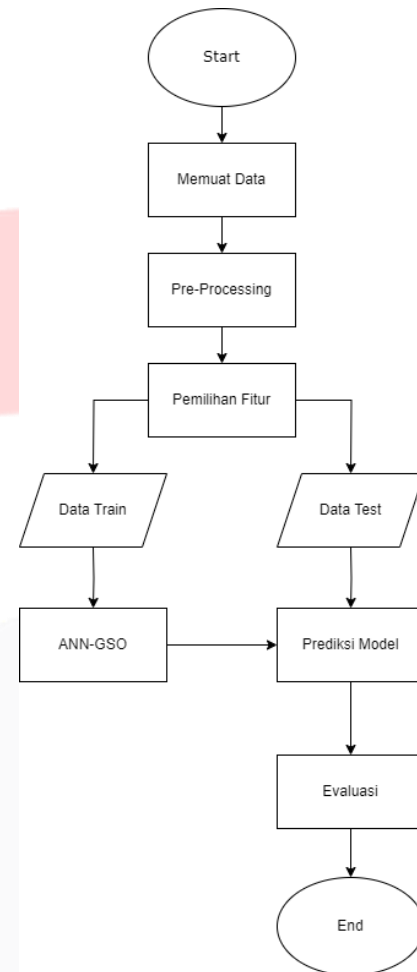
F1-Score menghitung rata rata antara *precision* dan *recall*. Rumus F1-Score seperti persamaan (3)

$$F1 - Score = 2 \frac{precision \times recall}{precision + recall} \quad (3)$$

III. SISTEM YANG DIBANGUN

Sistem yang akan dibuat adalah Prediksi Retweet Berdasarkan Fitur User-Based, Content-Based, Dan Time-Based Menggunakan Metode Klasifikasi ANN

Optimasi Glowworm Swarm Optimization. Berikut merupakan flowchart dari perancangan sistem yang akan dibangun.



GAMBAR 2.
Perancangan Sistem

Berdasarkan diagram alir pada Gambar 2, proses dimulai dengan memuat dataset, di mana dataset yang akan digunakan sudah tersedia, yaitu dataset U-20. Dataset U-20 merupakan data yang diambil dari Twitter dengan menggunakan kata kunci "Piala Dunia U-20". Setelah memuat data, kita akan melanjutkan ke tahap praproses, di mana pada tahap ini kita akan melakukan pembersihan data, seperti menghapus digit.

Setelah praproses selesai, langkah selanjutnya adalah pemilihan fitur yang akan digunakan, di mana fitur tersebut terbagi menjadi tiga kategori: user-based, time-based, dan content-based. Langkah berikutnya adalah pembagian data menjadi data latih dan data uji. Setelah data terbagi dengan baik, langkah selanjutnya adalah melakukan klasifikasi data menggunakan metode ANN-GSO. Setelah model klasifikasi dilatih, dilakukan evaluasi untuk mengukur kinerja data prediksi yang dihasilkan oleh model klasifikasi tersebut dengan menggunakan confusion matrix.

A. Dataset

Pada penelitian ini, dataset yang digunakan adalah dataset U-20 dengan jumlah 869 baris (entri data) dataset U-20 merupakan data yang diambil dari Twitter dengan kata kunci piala dunia usia 20 dan 28 kolom. Dalam dataset ini terdapat tiga fitur, yaitu content-based, time-based, dan user-based. Pengumpulan data dilakukan di Twitter, dan data diseleksi berdasarkan fitur-fitur tersebut. Data kemudian diberi label dan dibagi menjadi dua kelas: kelas 0 untuk tweet yang tidak di-retweet, dan kelas 1 untuk tweet yang di-retweet.

Dengan pembagian ini, kita dapat menganalisis lebih dalam faktor-faktor yang mempengaruhi kemungkinan suatu tweet di-retweet atau tidak. Sebagai contoh, dari dataset ini, kita menemukan bahwa ada 582 tweet yang tidak di-retweet dan 287 tweet yang di-retweet. Informasi ini dapat digunakan untuk memahami lebih lanjut pola interaksi pengguna dengan tweet-tweet tersebut. Oleh karena itu, harus dilakukan pemilihan fitur yang sesuai dengan data yang diharapkan.

B. Pemilihan Fitur

Pada tahap ini akan dilakukan pemilihan fitur yang dikategorikan berdasarkan content-based, time-based, dan user-based. Sehingga bisa dilihat pada tabel 2 fitur akhir yang didapatkan.

TABEL 2.
Fitur yang digunakan

Fitur yang digunakan				
Fitur		Nama	Tipe Data	Deskripsi
Content-Based	F1	has_hashtag	Boolean	Tweet mengandung hashtag
	F2	has_url	Boolean	Tweet mengandung URL
	F3	text_length	Numerik	Panjang teks pada tweet
	F4	has_image	Boolean	Tweet tersebut berisi gambar
	F5	has_video	Boolean	Tweet tersebut berisi video
Time-Based	F6	is_posted_in_noon	Boolean	Tweet ini dibuat dari jam 11.00-13.00
	F7	is_posted_in_eve	Boolean	Tweet ini dibuat dari jam 18.00-21.00
	F8	is_posted_on_weekend	Boolean	Tweet ini dibuat pada akhir pekan
User-Based	F9	Total_of_tweets	Numerik	Total tweet sebelumnya yang telah diposting pengguna di linimasa
	F10	No_of_favourite	Numerik	Jumlah tweet yang disukai pengguna di linimasa
	F11	No_of_followees	Numerik	Jumlah orang yang diikuti pengguna
	F12	No_of_follower	Numerik	Jumlah orang yang mengikuti pengguna
	F13	Age_of_account	Numerik	Jumlah hari sejak akun pengguna dibuat

IV. EVALUASI

A. Hasil Pengujian

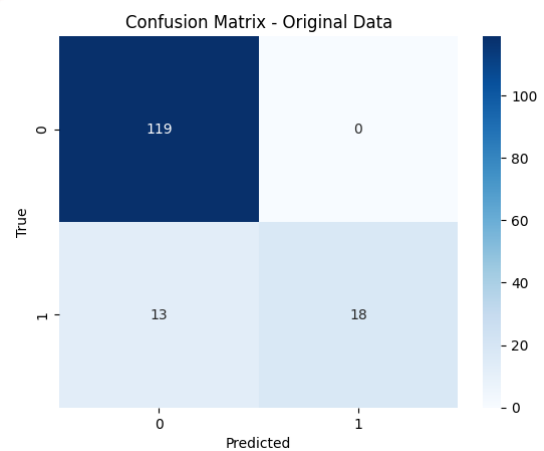
1. Hasil Skenario 1

Pada pengujian ini menggunakan ketidakseimbangan

dataset, yang bertujuan untuk mencari nilai akurasi, presisi, recall, dan F1-Score untuk dataset tersebut berdasarkan fitur *content-based*, *time-based*, dan *user-based*. dari pengujian ini diharapkan menemukan nilai default model (sebelum dioptimasi) dan best model (setelah dioptimasi) menggunakan algoritma Glowworm Swarm Optimization.

TABEL 3
Hasil Skenario 1

	Default Model	Best Model
Accuracy	68%	76%
Precision	70%	75%
Recall	68%	76%
F1-Score	69%	74%



GAMBAR 3.
Confusion Matrix - original data

Pada tabel 3 merupakan hasil pengujian skenario 1 untuk model ANN default model mendapatkan nilai akurasi 68%, presisi 70%, recall 68%, dan f1 skor 69%. Kemudian untuk best model mendapatkan nilai akurasi 76%, presisi 75%, recall 76%, dan f1 skor 74%.

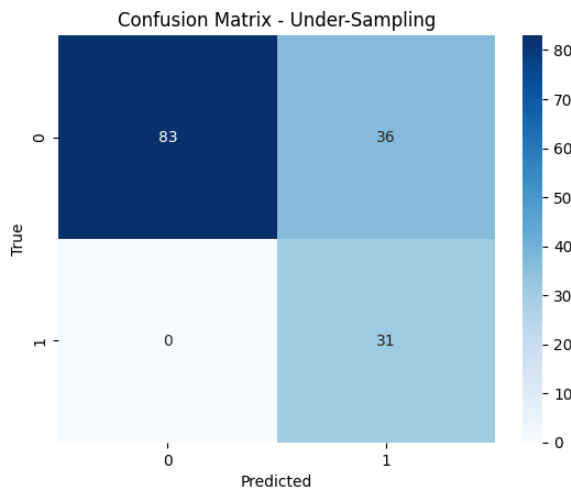
2. Hasil Skenario 2

Pada pengujian kedua ini, kelas-kelas yang tidak seimbang pada dataset ditangani menggunakan metode undersampling yang bertujuan untuk menemukan nilai default model (sebelum dioptimasi) dan best model (setelah dioptimasi) menggunakan algoritma Glowworm Swarm Optimization.

TABEL 4
Hasil Skenario 2

	Default Model	Best Model
Accuracy	60%	70%
Precision	64%	77%
Recall	60%	70%

F1-Score	61%	71%
----------	-----	-----



GAMBAR 4.
Confusion Matrix - Under Sampling

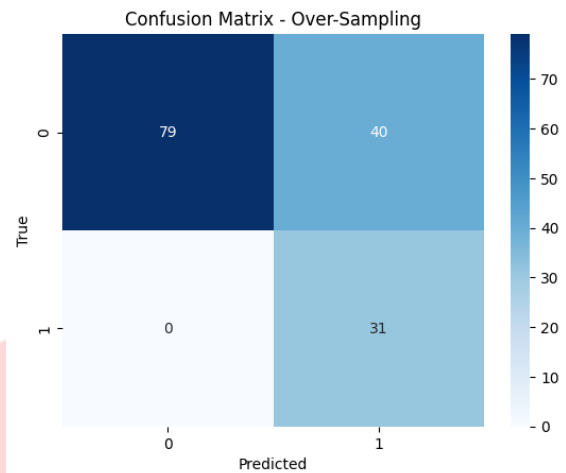
Pada tabel 4 merupakan hasil dari skenario 2 yang mendapatkan hasil ANN default model dengan nilai akurasi 60%, presisi 64%, recal 60%, dan f1 score 61%. Sedangkan untuk best model mendapatkan nilai akurasi 70%, presisi 77%, recall 70%, dan f1 score 71%.

3. Hasil Skenario 3

Pada pengujian 3 ini kelas-kelas yang tidak seimbang pada dataset kembali digunakan dan ditangani dengan menggunakan metode oversampling yang bertujuan menemukan nilai default model (sebelum dioptimasi) dan best model (setelah dioptimasi) menggunakan algoritma Glowworm Swarm Optimization

TABEL 5
Hasil skenario 3

	Default Model	Best Model
Accuracy	58%	78%
Precision	66%	79%
Recall	58%	78%
F1-Score	60%	78%



GAMBAR 5.
Confusion Matrix - Over Sampling

Pada tabel 5 merupakan hasil dari skenario 3 yang mendapatkan hasil ANN default model dengan nilai akurasi 58%, presisi 66%, recal 58%, dan f1 score 60%. Sedangkan untuk best model mendapatkan nilai akurasi 78%, presisi 79%, recall 78%, dan f1 score 78%.

B. Analisis Hasil Pengujian

Hasil pengujian menunjukkan bahwa penerapan Glowworm Swarm Optimization (GSO) secara keseluruhan meningkatkan performa model pada berbagai skenario dataset. Pada skenario ketidakseimbangan dataset, undersampling, dan oversampling, model yang dioptimasi dengan GSO menunjukkan peningkatan signifikan dalam semua metrik performa dibandingkan dengan model default yang memiliki akurasi sebesar 78% dan F1-Score 78%. Ini mengindikasikan bahwa GSO sangat efektif dalam mengatasi ketidakseimbangan dataset, dengan oversampling sebagai metode yang paling menguntungkan setelah optimasi. Meskipun undersampling juga meningkatkan performa secara signifikan, hasil optimasi pada dataset yang tidak seimbang menunjukkan pentingnya teknik ini dalam meningkatkan akurasi dan keandalan model secara keseluruhan.

V. KESIMPULAN

Penelitian ini berhasil mengembangkan model prediksi retweet dengan menggunakan metode ANN-GSO. Hasil penelitian menunjukkan bahwa penerapan GSO terdapat peningkatan performa model pada berbagai skenario dataset. Untuk hasil default model memiliki nilai akurasi 58% dan nilai f1 score 60%, sedangkan untuk hasil ANN-GSO memiliki nilai akurasi 78% dan nilai f1 score 78% hal ini menunjukkan adanya peningkatan performa.

REFERENSI

- [1] M. Rizqi Akbar, "Prediksi Retweet Berdasarkan User-Based dan Content-Based Menggunakan Metode Ensemble Stacking," 2023.
- [2] V. E. Candrika, J. Jondri, and I. Indwiarti, "Retweet Predictions Regarding COVID-19 Vaccination Tweets through The Method of Multi Level Stacking," *JINAV: Journal of Information and Visualization*, vol. 4, no. 1, pp. 28–35, Jan. 2023, doi: 10.35877/454RIJINAV1518.
- [3] B. M. Pasaribu, Y. Winoto, U. Lies, and S. Khadijah, "Twitter Sebagai Media Pemenuhan Kebutuhan Informasi Penggemar Musik Korea di Indonesia," *Ilmu Informasi Perpustakaan dan Kearsipan*, vol. 11, no. 1, pp. 7–15, Aug. 2022, doi: 10.24036/117024-0934.
- [4] E. Penggunaan *et al.*, "Efektivitas Penggunaan Twitter Sebagai Sarana Peningkatan Berpikir Kritis Mahasiswa Ilmu Komunikasi," *MUKASI: Jurnal Ilmu Komunikasi*, vol. 2, no. 1, pp. 18–28, Feb. 2023, doi: 10.54259/MUKASI.V2I1.1346.
- [5] E. T. Arifin, J. Jondri, and I. Indwiarti, "Prediction Retweet Using User-Based and Content-Based with ANN-GA Classification Method," *Building of Informatics, Technology and Science (BITS)*, vol. 4, no. 2, pp. 522–528, Sep. 2022, doi: 10.47065/BITS.V4I2.1931.
- [6] Y. Setiawan, J. Jondri, and W. Astuti, "Twitter Sentiment Analysis on Online Transportation in Indonesia Using Ensemble Stacking," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 6, no. 3, pp. 1452–1458, Jul. 2022, doi: 10.30865/MIB.V6I3.4359.
- [7] A. R. A. Maden, J. Jondri, and W. Astuti, "Analysis of Community Sentiment on Twitter towards COVID-19 Vaccine Booster Using Ensemble Bagging Methods," *Building of Informatics, Technology and Science (BITS)*, vol. 4, no. 2, pp. 554–561, Sep. 2022, doi: 10.47065/BITS.V4I2.1973.
- [8] D. Puspendari, C. Author, and J. Telekomunikasi Terusan Buah Batu Bandung, "Retweet Prediction Using Multi-Layer Perceptron Optimized by The Swarm Intelligence Algorithm," *Jurnal Online Informatika*, vol. 8, no. 2, pp. 252–260, Dec. 2023, doi: 10.15575/JOIN.V8I2.1193.
- [9] D. U. Suhendra, J. Jondri, and I. Indwiarti, "Sentiment Analysis of Hate Speech on Twitter Public Figures with AdaBoost and XGBoost Methods," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 6, no. 3, pp. 1484–1491, Jul. 2022, doi: 10.30865/MIB.V6I3.4394.
- [10] R. Artikel, L. Surya Haryadi, B. Renaldy Suteja, and U. Kristen Maranatha Jl Profdrgr Suria Sumantri No, "Prediksi Penyebaran Informasi di Twitter dengan Metode Pembelajaran Mesin dengan Fitur Linimasa," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 7, no. 1, pp. 2443–2229, Apr. 2021, doi: 10.28932/JUTISI.V7I1.3324.
- [11] M. Iqbal Zakasih, W. Tri Handoko, and J. Tri Lomba Juang No, "ANALISIS SENTIMEN PENGGUNA TWITTER TENTANG NFT (NON FUNGIBLE TOKEN) DENGAN METODE NAIVE BAYES CLASSIFIER," *Jurnal Informatika dan Rekayasa Elektronik*, vol. 5, no. 2, pp. 221–229, Nov. 2022, doi: 10.36595/JIRE.V5I2.694.
- [12] F. Solihin, S. Awaliyah, A. Muid, and A. Shofa, "Pemanfaatan Twitter Sebagai Media Penyebaran Informasi Oleh Dinas Komunikasi dan Informatika," *Jurnal Pendidikan Ilmu Pengetahuan Sosial*, vol. 13, no. 1, pp. 52–58, Jun. 2021, Accessed: Sep. 01, 2024. [Online]. Available: <https://e-journal.upr.ac.id/index.php/JIP-IPS/article/view/2813>
- [13] R. A. P. Nugroho, "Penggunaan Tone Indicator dalam Pencegahan Miskomunikasi di Media Sosial Twitter/X," 2024, Accessed: Sep. 01, 2024. [Online]. Available: <https://dspace.uui.ac.id/handle/123456789/50529>
- [14] F. F. Noorikhshan, H. Ramdhani, B. C. Sirait, and N. Khoerunisa, "Dinamika Internet, Media Sosial, dan Politik di Era Kontemporer: Tinjauan Relasi Negara-Masyarakat," *Journal of Political Issues*, vol. 5, no. 1, pp. 95–109, Jul. 2023, doi: 10.33019/JPI.V5I1.131.
- [15] R. H. Dananjaya, S. Sutrisno, and S. Fitriady, "PENERAPAN ARTIFICIAL NEURAL NETWORK (ANN) DALAM MEMPREDIKSI KAPASITAS DUKUNG FONDASI TIANG," *Matriks Teknik Sipil*, vol. 10, no. 4, pp. 419–426, Dec. 2022, doi: 10.20961/MATEKSI.V10I4.65034.
- [16] F. Amir, E. Utami, and H. Hanafi, "Literature Study on the Development of Neural Networks For Weather Forecasting," *J Teknol*, vol. 17, no. 1, pp. 49–57, Jun. 2024, doi: 10.34151/JURTEK.V17I1.4637.
- [17] E. AGUSRIADI, "Sistem Pakar dalam Menganalisis Penyakit Organ dan Jaringan Tubuh dengan Metode Perceptron dan Fitur Augmented Reality," 2022.