

Prediksi Retweet Menggunakan Fitur Content Based dengan Metode Jaringan Saraf Tiruan Dioptimasi dengan Particle Swarm Optimization

Alifia Belqis¹, Jondri², Widi Astuti³

^{1,2,3}Fakultas Informatika, Universitas Telkom, Bandung

¹alifiabilqis@students.telkomuniversity.ac.id, ²jondri@telkomuniversity.ac.id, ³widiwdu@telkomuniversity.ac.id

Abstrak

Twitter merupakan platform media sosial yang memiliki jutaan pengguna di seluruh dunia, dengan tweet semua orang atau pengguna dapat melihat informasi secara *up to date* dan dengan fitur *trending* yang biasanya menggunakan *hashtag*. *Tweet* dapat dibagikan ke pengguna lainnya dengan cara *retweet* yang sangat berperan penting dalam penyebaran informasi yang lebih luas. Penelitian ini membahas mengenai prediksi retweet menggunakan fitur berbasis konten dengan metode jaringan saraf tiruan untuk klasifikasinya dan dioptimalkan dengan algoritma *Particle Swarm Optimization* untuk mendapatkan tingkat akurasi yang lebih tinggi. Model Jaringan Syaraf Tiruan dengan *Particle Swarm Optimization* ini menunjukkan hasil terbaik ketika sudah dilakukan skenario *oversampling* dengan hasil akurasi 88.66% dan *F1-score* 89.19%.

Kata kunci : *Twitter, JST-PSO, retweet, content-based, F1-Score*

Abstract

Twitter is a social media platform that has millions of users worldwide. With tweets, everyone or users can see information that is up to date and with trending features that usually use hashtags. Tweets can be shared with other users by retweeting which plays a very important role in the wider dissemination of information. This study discusses retweet prediction using content-based features using the Neural Network method for classification and optimizing it with the Particle Swarm Optimization algorithm to obtain a higher level of accuracy. This Artificial Neural Network Model with Particle Swarm Optimization shows the best results when the oversampling scenario has been carried out with an accuracy of 88.66% and an F1-score of 89.19%.

Keywords: *Twitter, JST-PSO, retweet, content-based, F1 Score*

1. Pendahuluan

Latar Belakang

Latar belakang penelitian ini mengacu pada pemilihan Twitter sebagai subjek utama analisis. Twitter merupakan salah satu platform media sosial yang telah mencapai popularitas yang besar dalam era digital saat ini. Twitter memungkinkan penggunanya untuk menulis dan membagikan pendapat mereka melalui tulisan pribadi atau konten yang dibagikan oleh orang lain. Menurut data dari Statistika, pada kuartal II-2021, Twitter memiliki 206 juta pengguna aktif harian di seluruh dunia.[1]. Selain itu pada media sosial Twitter juga terdapat fitur *Retweet*. *Retweet* merupakan fitur yang memposting ulang sebuah *tweet* baik *tweet* diri sendiri maupun *tweet* orang lain yang mengartikan kita juga menyukai atau setuju dengan *tweet* tersebut. Postingan/*tweet* yang diretweet biasanya bertujuan untuk mempublikasikan lebih luas lagi agar *tweet* tersebut dapat dijangkau oleh lebih banyak orang.

Twitter memberikan keterbatasan dalam penulisan, dengan batasan maksimal sebanyak 140 karakter dalam setiap posting (*tweet*). Pengguna Twitter, baik yang terdaftar maupun yang bukan pengguna, dapat memantau perkembangan topik tertentu dengan mengikuti tagar (*hashtag*) yang ditandai dengan simbol tanda pagar (#). Tanda pagar ini memungkinkan individu untuk mengikuti serta berpartisipasi dalam diskusi seputar topik tertentu yang menarik perhatian.[2]. Maka dari itu, platform twitter memiliki fitur di mana seseorang dapat menemukan informasi dari suatu kata kunci yang sedang banyak dibicarakan oleh banyak pengguna twitter.

Dalam penelitian ini, peneliti mengembangkan model dengan menggunakan 16.000 data *tweet* yang diperoleh melalui perpustakaan Tweepy. Kemudian, model ini menggunakan dua jenis fitur, yaitu fitur yang berkaitan dengan pengguna dan fitur yang berkaitan dengan konten. Data *tweet* yang digunakan dalam penelitian ini ditulis dalam bahasa Indonesia, dan tujuan penelitian ini adalah membangun model yang dapat memprediksi dan mengukur keakuratan apakah suatu konten akan di-retweet atau tidak.

Model prediksi retweet ini memanfaatkan Jaringan Saraf Tiruan yang telah dioptimalkan menggunakan Particle Swarm Optimization (PSO). PSO sendiri adalah sebuah algoritma berbasis populasi yang mengoptimalkan individu-individu dalam proses pencarian. Dalam PSO, populasi disebut sebagai "swarm" dan individu-individu disebut sebagai "particle." Setiap partikel bergerak dengan kecepatan yang disesuaikan berdasarkan informasi dari lingkungan pencarian, dan mereka menyimpan catatan dari posisi terbaik yang pernah mereka capai. [3].

Topik dan Batasannya

Topik pada penelitian ini adalah melakukan pengujian model jaringan syaraf tiruan yang dioptimasi *Particle Swarm Optimization*. Penelitian ini memiliki batasan masalah yaitu dataset yang digunakan merupakan *dataset* Bahasa Indonesia dengan menggunakan kata kunci *blackpink*. Dataset tersebut terdiri dari 16000 data tweet terhitung dari bulan Maret-Juni 2023.

Tujuan

Tujuan penelitian ini adalah untuk melihat performansi dari model jaringan syaraf tiruan yang dioptimasi *Particle Swarm Optimization* untuk melakukan prediksi *retweet*. Skenario pengujian yang dilakukan pada penelitian yaitu ada dua skenario, skenario pertama menggunakan *Random Under Sampling* dan skenario kedua menggunakan *Random Over Sampling*. Langkah selanjutnya adalah menghitung akurasi dan F1-score dari model JST-PSO dan melakukan perbandingan model JST saja dengan JST-PSO.

Organisasi Tulisan

Setelah bagian pendahuluan dalam Bab 1, tahapan berikutnya pada Bab 2 akan mengulas studi literatur yang relevan dengan kegiatan penelitian yang direncanakan. Pada Bab 3, akan dibahas teori dan sistem yang diimplementasikan dalam penelitian ini. Bab 4 akan membahas hasil analisis dari penelitian yang telah dilaksanakan, diikuti oleh penjelasan kesimpulan yang diperoleh berdasarkan temuan penelitian pada Bab 4. Kesimpulan tersebut akan dijelaskan lebih lanjut dalam Bab 5.

2. Studi Terkait

Pada penelitian sebelumnya dengan judul “Prediksi Retweet Menggunakan Fitur Berbasis Pengguna dan Fitur Berbasis Konten dengan Metode Klasifikasi Artificial Neural Network (ANN),” yang ditulis oleh Hamidan Amarullah Purwaatmaja di tahun 2021 ini menggunakan model *K-Fold Cross Validation*, *Artificial Neural Network (ANN)*, *Backpropagation*, *Confusion Matrix*, *NearMiss* yang bertujuan untuk membangun model prediksi *retweet* dengan data yang diambil menggunakan twitter API, Fitur yang dimanfaatkan meliputi fitur yang berbasis pengguna dan fitur yang berbasis konten. Pemanfaatan fitur-fitur tersebut berpotensi memengaruhi fenomena *retweet*. serta model terbaik yang dapat mengatasi imbalance class adalah metode *undersampling* [4].

Pada tahun 2021, Muhammad Syah Zannuar Suwito melakukan penelitian yang berjudul “Prediksi Retweet Berdasarkan Feature User-Based Menggunakan Metode Klasifikasi Random Forest”. Maksud dan tujuan dari penelitian ini adalah melakukan prediksi terhadap sebuah *tweet* yang akan mendapatkan *retweet* atau tidaknya dengan metode klasifikasi *Random Forest* dengan berfokus terhadap *feature User-Based*. Dalam penelitian ini menghasilkan performansi yang cukup baik dengan menggunakan *fold cross validation*, juga penerapan dari *hyperparameter tuning* nilai performansi pun meningkat [5].

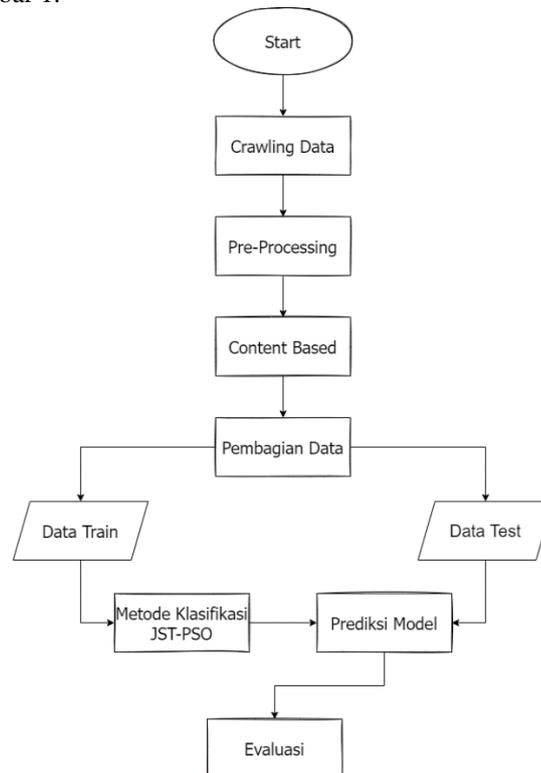
Pada tahun 2021, Agung Gumelar melakukan penelitian yang berjudul “Prediksi Retweet Berdasarkan Fitur User-Based menggunakan Metode Klasifikasi Naïve Bayes”. Tujuan dari penelitian ini adalah memprediksi apakah suatu tweet akan menjadi populer dengan *User-Based Feature*. *Dataset* tersebut akan diuji menggunakan model Naïve Bayes untuk mengetahui apakah model ini cocok untuk diaplikasikan ke *dataset* ini. Adapun hasil dari penelitian ini adalah model *Naïve Bayes* yang cukup bagus dalam memprediksi tweet yang akan menjadi populer atau tidak. Hasil dari *undersampling* dan *oversampling* pun tidak berpengaruh banyak terhadap hasil klasifikasi [6].

Pada tahun 2023, Muhammad Rizqi Akbar mengadakan penelitian berjudul "Prediksi Retweet Berdasarkan Pendekatan User-Based dan Content-Based Menggunakan Metode Ensemble Stacking." Penelitian ini fokus pada pembuatan model untuk meramalkan *retweet* berdasarkan fitur user-based dan content-based menggunakan teknik Ensemble Stacking melalui proses validasi silang K-fold. Ensemble Stacking ini terdiri dari tiga base-learner, yaitu Random Forest, Gradient Boosting, dan Support Vector Machine (SVM), sementara meta-learner yang digunakan adalah Support Vector Machine (SVM). Hasil pemodelan ini menunjukkan hasil terbaik saat dilakukan penanganan ketidakseimbangan kelas (Imbalanced Class Handling) dengan menggunakan Teknik SMOTE dan validasi silang K-fold dengan nilai k=10, yang menghasilkan F1-score sebesar 86.46%. Berdasarkan hasil ini, dapat disimpulkan bahwa model yang telah dibangun berhasil meningkatkan kinerja prediksi dari base-learner-nya.[7].

Pada tahun 2021, Rakes melakukan sebuah studi yang berjudul "Prediksi Retweet Menggunakan Fitur Berbasis Pengguna dengan Metode Klasifikasi Support Vector Machine." Tujuan dari penelitian ini adalah menciptakan sebuah program yang dapat memprediksi apakah suatu tweet akan mendapatkan *retweet* atau tidak, serta mengevaluasi kinerja dan akurasi mesin pembelajaran Support Vector Machine dengan menggunakan fitur berbasis pengguna sebagai atribut. Setelah melakukan preprocessing data, penelitian ini menerapkan metode k-fold cross validation. Hasil penelitian ini menghasilkan algoritma yang mampu memprediksi kemungkinan *retweet* sebuah tweet dengan F1-score sebesar 66,05%.[8].

3. Sistem yang Dibangun

Sistem yang dikembangkan untuk penelitian ini didesain dengan menerapkan metode JST-PSO, yang dapat dilihat dalam ilustrasi pada Gambar 1.



Gambar 1. Rancangan sistem prediksi retweet menggunakan klasifikasi JST-PSO

Berdasarkan diagram alir yang ditampilkan pada Gambar 1, proses dimulai dengan melakukan crawling data atau pengambilan data secara manual menggunakan *snsrape* dan mempersiapkan dataset yang siap untuk diolah. Setelah itu, dilanjutkan dengan tahap *preprocessing*. Pada tahap ini, dilakukan beberapa langkah seperti, pembersihan data dengan menghapus URL, mention, emoji, digit, dan tanda baca, melakukan *tokenisasi* untuk memecah kalimat menjadi kata-kata, normalisasi data dengan menggunakan metode *Min-Max Scaling*.

Setelah proses *preprocessing* selesai, langkah selanjutnya adalah pembagian data menjadi data *train* dan data *test*. Setelah data terbagi dengan baik, langkah terakhir dari proses ini adalah klasifikasi data dengan menggunakan metode jaringan syaraf tiruan – *Particle Swarm Optimization*. Setelah model klasifikasi dilatih, dilakukan evaluasi untuk mengukur kinerja data prediksi yang dihasilkan oleh model klasifikasi tersebut menggunakan *confusion matrix*.

3.1 Pengambilan dan Pengumpulan *Dataset*

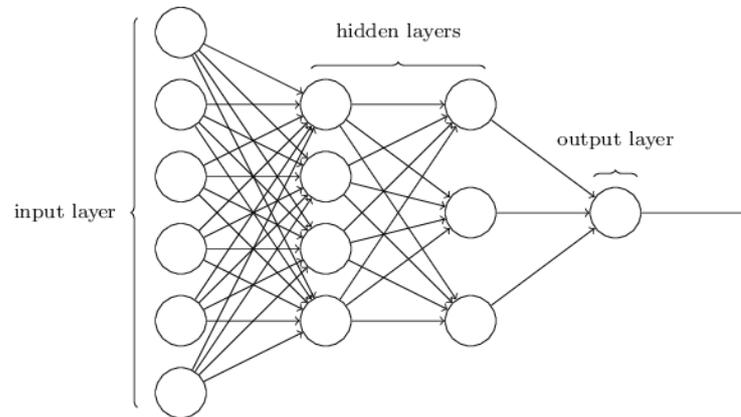
Pengambilan data dilakukan melalui *Twitter API* yang dijalankan dalam bahasa pemrograman Python. Melalui proses pengambilan data ini, kami memperoleh kumpulan data yang disebut sebagai dataset. Untuk melakukan pengambilan data ini, kami mengandalkan sebuah perpustakaan yang dikenal sebagai *Tweepy*. *Tweepy* adalah sebuah pustaka *Python* yang digunakan untuk mengakses *API Twitter*. Dengan bantuan *Tweepy*, kita dapat mengambil informasi dari *Twitter* menggunakan skrip Python. [9]. Dalam konteks penelitian ini, *snsrape* digunakan untuk mengumpulkan data dan informasi yang terkait dengan kata kunci "*blackpink*" yang sedang beredar di *platform Twitter*.

3.2 Jaringan Saraf Tiruan (*Artificial Neural Network*)

Neural Network atau jaringan saraf tiruan adalah salah satu representasi buatan yang selalu mencoba mensimulasikan proses pembelajaran pada otak manusia [10]. Jaringan saraf tiruan terdiri atas lapisan input dan lapisan output. Tiap lapisan memiliki beberapa unit neuron yang mempunyai sebuah fungsi aktivasi yang menentukan dari unit keluar [11].

Dalam struktur Jaringan Saraf Tiruan, terdapat beberapa lapisan yang terlibat. Pertama, terdapat lapisan input yang terdiri dari unit-unit yang menerima pola masukan data dari luar, yang mencerminkan informasi terkait masalah yang sedang dihadapi. Kemudian, terdapat lapisan *hidden* yang terdiri dari unit-unit yang mengolah informasi secara internal dan hasilnya tidak dapat diamati secara langsung. Terakhir, ada lapisan output yang terdiri dari unit-unit yang menghasilkan solusi yang memecahkan masalah yang diberikan [12].

Hubungan antar neuron atau yang biasa disebut sebagai arsitektur jaringan. Neuron-neuron tersebut terkumpul dalam lapisan-lapisan yang disebut layer, lapisan tersebut dibagi menjadi tiga bagian yaitu *input layer*, *hidden layer*, *output layer* [13] seperti pada Gambar 2 [14].



Gambar 2. Arsitektur Jaringan Syaraf Tiruan

3.3 Particle Swarm Optimization

Particle Swarm Optimization (PSO) adalah sebuah metode optimisasi heuristik global yang pertama kali diperkenalkan oleh Dokter Kennedy dan Eberhart pada tahun 1995. Metode ini terinspirasi oleh perilaku sosial kawanan burung yang berusaha mencapai tujuan yang tidak mereka ketahui [15].

Metode PSO adalah teknik optimasi di mana pencarian solusi mengikuti perilaku sosial yang terlihat dalam populasi burung (*flock of birds*) dan populasi ikan (*school of fish*) saat mereka berusaha bertahan hidup [16]. Sebagai contoh, pada kelompok ikan yang sedang bermigrasi, ikan mampu mengetahui jarak satu sama lain berkat gelombang air yang membantu mereka. Dampak dari hal ini adalah ikan dapat bergerak dengan cepat dan menjaga jarak yang akurat dalam kelompok mereka [17].

3.4 Confusion Matrix

Evaluasi dan validasi adalah suatu proses akhir untuk mengetahui kinerja sistem prediksi yang dibuat. Maka dari itu penulis melakukan pengukuran kinerja model prediksi retweet menggunakan confusion matrix. Confusion matrix adalah tabel yang menyatakan klasifikasi jumlah data uji yang benar dan jumlah data uji yang salah [18]. Contoh confusion matrix untuk klasifikasi biner seperti pada Tabel 1.

Tabel 1. Confusion Matrix

	Kelas Prediksi	
	1	0
Kelas	1 TP	0 FN
Aktual	0 FP	TN

Adapun keterangan dari Tabel 1 sebagai berikut [19]:

TP (True Positive) = jumlah dokumen dari kelas 1 yang benar diklasifikasikan sebagai kelas 1.

TN (True Negative) = jumlah dokumen dari kelas 0 yang benar diklasifikasikan sebagai kelas 0.

FP (False Positive) = jumlah dokumen dari kelas 0 yang salah diklasifikasikan sebagai kelas 1.

FN (False Negative) = jumlah dokumen dari kelas 1 yang salah diklasifikasikan sebagai kelas 0.

Rumus untuk menghitung *confusion matrix* dari *accuracy*, *precision*, *recall*, dan *F1-score* adalah :

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$F1\ Score = 2 \frac{Recall \times Precision}{Recall + Precision} \quad (4)$$

4. Evaluasi

Pembahasan ini menjelaskan lebih detail hasil penerapan model Jaringan Syaraf Tiruan dengan dioptimalkan menggunakan Particle Swarm Optimization. Pengujian dilakukan dengan membandingkan hasil pembobotan dari pengujian model Jaringan Syaraf Tiruan saja dengan Jaringan Syaraf Tiruan yang dioptimalkan dengan Particle Swarm Optimization untuk mendapatkan hasil perbandingan akurasi antara JST dan JST-PSO dari model prediksi retweet.

4.1 Hasil Pengujian

4.1.1 Dataset

Pengumpulan data dilakukan dengan teknik scraping twitter menggunakan snsrape. Hasil dari scraping ini mendapatkan 16.000 tweets Bahasa Indonesia dengan kata kunci pencarian “blackpink” dengan rentang waktu bulan Maret - Juni 2023.

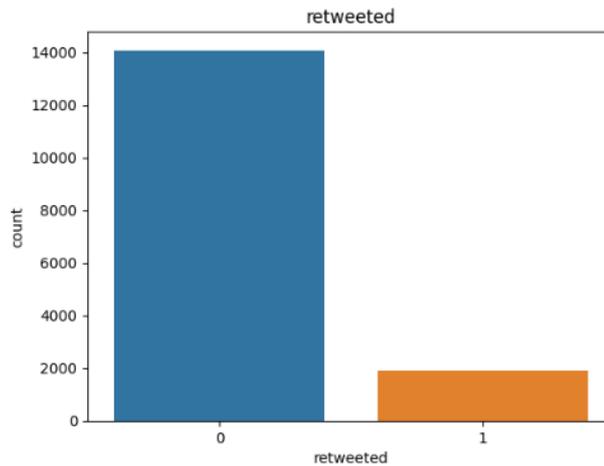
Data yang diperoleh dari hasil crawling data tweet ini memiliki karakteristik berdasarkan pengguna dan berdasarkan konten. Selanjutnya, dilakukan pemilihan fitur-fitur yang kurang diperlukan untuk proses klasifikasi. Tabel 2 menunjukkan contoh dataset yang digunakan beserta fitur-fiturnya.

Table 2. Fitur yang digunakan

Fitur	Nama	Tipe Data
F1	total_tweet	Int64
F2	no_follower	Int64
F3	no_following	Int64
F4	age_of_account	Int64
F5	no_of_favorite	Int64
F6	avg_tweet_per_day	Float64
F7	avg_fav_per_day	Float64
F8	verified	Int64
F9	len_of_text	Int64
F10	opt_length	Int64
F11	contain_picture	Int64
F12	contain_video	Int64
F13	contain_upper	Int64
F14	contain_hashtag	Int64
F15	con_user_mentioned	Int64
F16	retweeted	Int64
F17	sen_negatif	Int64
F18	sen_netral	Int64
F19	sen_positif	Int64

4.1.2 Evaluasi Model

Data yang diperoleh dari proses crawling diawali dengan proses pengolahan untuk memastikan bahwa data tersebut dapat digunakan dalam model. Selanjutnya, dilakukan pengecekan terhadap ketidakseimbangan kelas pada data yang telah diperoleh. Kelas 0 menandakan tweet yang tidak di-retweet, sementara kelas 1 menandakan tweet yang di-retweet. Terdapat ketidakseimbangan jumlah data antara kedua kelas, di mana kelas 0 memiliki 14077 data, sedangkan kelas 1 hanya memiliki 1923 data. Gambar 3 menggambarkan visualisasi hasil dari pengecekan ketidakseimbangan ini.

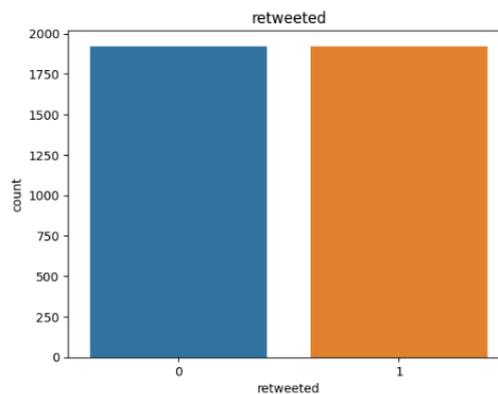


Gambar 3. Visualisasi Imbalance Class

Untuk mengatasi permasalahan imbalance class ini, dilakukan proses handling dengan dua skenario pengujian yang berbeda. Skenario pertama menggunakan metode undersampling dengan teknik random undersampling, sementara skenario kedua menggunakan metode oversampling dengan teknik random oversampling. Kedua skenario penanganan ketidakseimbangan ini akan dibahas lebih lanjut pada bagian berikutnya.

4.1.3 Skenario Pertama

Pada pengujian ini, dataset yang digunakan adalah dataset yang sudah dilakukan Imbalanced Class Handling menggunakan RandomUnderSampler. Visualisasi data dari hasil random undersampling seperti pada Gambar 4.



Gambar 4. Visualisasi Hasil Random Under Sampling

Pada pengujian pertama, hasil yang didapatkan untuk melakukan prediksi retweet menggunakan metode Jaringan Syaraf Tiruan ini akurasi rata-rata dari 5-fold yang digunakan adalah 70.93%, sedangkan untuk F1-score mendapatkan 70.58%. Tabel 3 menunjukkan hasil pengujian dengan Jaringan Syaraf Tiruan menggunakan RandomUnderSampling.

Table 3. Hasil Model Jaringan Syaraf Tiruan Random Under Sampling

Fold	Akurasi	Presisi	Recall	F1 Score
1	72.72%	75.81%	66.75%	70.99%
2	71.26%	70.42%	73.17%	71.77%
3	70.10%	71.38%	66.92%	69.08%
4	73.86%	73%	75.84%	74.39%
5	66.71%	66.84%	66.49%	66.66%
Mean	70.93%	71.49%	69.83%	70.58%

Hasil yang didapatkan dari menggunakan metode Jaringan Syaraf Tiruan yang dioptimasi Particle Swarm Optimization menghasilkan akurasi dengan rata-rata dari 5-fold adalah 72.23%, sedangkan untuk F1-score nya mendapatkan rata-rata 73.77%. Pada Tabel 4 ini menunjukkan hasil pengujian dari JST-PSO menggunakan RandomUnderSampling.

Table 4. Hasil Model JST-PSO Random Under Sampling

Fold	Akurasi	Presisi	Recall	F1 Score
1	74.41%	77.01%	63.11%	76.47%
2	72.04%	69.77%	78.38%	75.24%
3	70.74%	66.74%	58.85%	70.21%
4	74.12%	71.64%	84.15%	70.29%
5	69.83%	69.03%	73.50%	73.77%
Mean	72.23%	70.84%	73.50%	73.77%

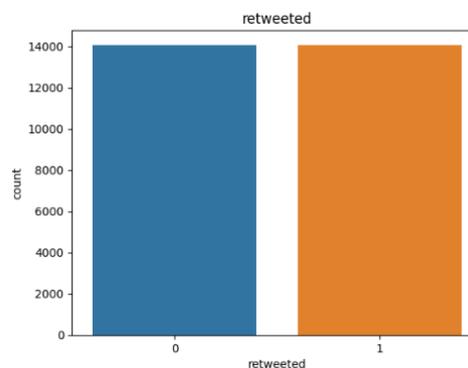
Dapat dilihat perbandingan dari hasil klasifikasi Jaringan Syaraf Tiruan dan Jaringan Syaraf Tiruan dengan Particle Swarm Optimization. Hasil akurasi tertinggi didapatkan pada model JST-PSO dengan rata-rata 72.23%, sedangkan hasil F1-score tertinggi didapatkan dengan metode JST-PSO dengan 73.77%. Visualisasi perbandingan hasil rata-rata dari model JST dengan JST-PSO dapat dilihat pada Tabel 5.

Table 5. Hasil Model JST-PSO Random Under Sampling

Model	Akurasi	Presisi	Recall	F1 Score
JST	70.93%	71.49%	69.83%	70.58%
JST-PSO	72.23%	70.84%	73.50%	73.77%

4.1.4 Skenario Kedua

Pada pengujian ini, dataset yang digunakan adalah dataset yang sudah dilakukan Imbalanced Class Handling menggunakan RandomOverSampler. Dengan tujuan meningkatkan sampel kelas minoritas sampai sama dengan kelas mayoritas lain dengan menduplikasi secara acak sampel kelas minoritas. Visualisasi data dari hasil random oversampling seperti pada Gambar 5.



Gambar 5. Visualisasi Hasil Random Over Sampling

Pada pengujian kedua, dilakukan Oversampling dengan menggunakan metode random oversampling pada dataset, sehingga jumlah dataset sekarang berjumlah 28154 data. Hasil dari penjumlahan dataset yang cukup signifikan dibandingkan dari skenario pertama, metode Jaringan Syaraf Tiruan ini mendapatkan nilai rata-rata dari 5-fold akurasi 85.74% dan F1-score 86.43%. Pada Tabel 6 menunjukkan hasil dari 5-fold dan rata model Jaringan Syaraf Tiruan dengan RandomOverSampling.

Table 6. Hasil Model Jaringan Syaraf Tiruan Random Over Sampling

Fold	Akurasi	Presisi	Recall	F1 Score
1	85.15%	85.49%	88.21%	85.16%
2	85.01%	81.97%	84.65%	85.48%
3	84.54%	80.45%	91.29%	85.38%
4	87.32%	85.88%	89.62%	88.48%
5	86.69%	80.95%	91.08%	87.64%
Mean	85.74%	82.95%	88.97%	86.43%

Hasil yang didapatkan dari menggunakan metode Jaringan Syaraf Tiruan dengan Particle Swarm Optimization menghasilkan akurasi dengan rata-rata dari 5-fold adalah 88.66%, sedangkan untuk F1-score nya mendapatkan rata-rata 89.19%. Pada Tabel 7 ini menunjukkan hasil pengujian dari JST-PSO menggunakan RandomOverSampling.

Table 7. Hasil Model JST-PSO Random Over Sampling

Fold	Akurasi	Presisi	Recall	F1 Score
1	88.40%	83.34%	95.98%	89.22%
2	88.52%	84.42%	94.49%	89.17%
3	87.63%	85.70%	90.33%	87.96%
4	90.21%	88.45%	92.50%	90.43%
5	88.54%	84.44%	94.49%	89.18%
Mean	88.66%	85.27%	93.56%	89.19%

Dapat dilihat bahwa perbandingan dari hasil klasifikasi Jaringan Syaraf Tiruan dan Jaringan Syaraf Tiruan dengan Particle Swarm Optimization pada scenario 2. Hasil akurasi tertinggi didapatkan pada model JST-PSO dengan rata-rata 88.6%, sedangkan hasil F1-score tertinggi didapatkan dengan metode JST-PSO dengan 89.19%. Visualisasi perbandingan hasil rata-rata dari model JST dengan JST-PSO dapat dilihat pada Tabel 8.

Table 8. Hasil Model JST-PSO Random Over Sampling

Model	Akurasi	Presisi	Recall	F1 Score
JST	85.74%	82.95%	88.97%	86.43%
JST-PSO	88.66%	85.27%	93.56%	89.19%

4.2 Analisis Hasil Pengujian

Hasil penelitian prediksi retweet menggunakan metode Jaringan Syaraf Tiruan dan optimalisasi Particle Swarm Optimization menunjukkan bahwa akurasi tertinggi dicapai saat penanganan kelas yang tidak seimbang diterapkan pada skenario kedua menggunakan RandomOverSampling, dengan tingkat akurasi mencapai 88.66%. Dari hasil kedua skenario yang dijalankan, dapat disimpulkan bahwa model Jaringan Syaraf Tiruan yang telah dioptimasi dengan Particle Swarm Optimization memberikan hasil yang lebih baik dalam hal akurasi dan F1-score dibandingkan dengan menggunakan model Jaringan Syaraf Tiruan tanpa optimalisasi.

5. Kesimpulan

Berdasarkan penelitian yang telah dilakukan, kesimpulannya adalah bahwa penggunaan metode Jaringan Syaraf Tiruan yang dioptimalkan dengan Particle Swarm Optimization untuk memprediksi retweet berdasarkan fitur user-based dan content-based menghasilkan peningkatan dalam kinerja, terutama dari segi akurasi dan F1-score. Dalam penelitian ini, dua skenario pengujian dilakukan, yaitu pengujian klasifikasi dengan undersampling dan pengujian klasifikasi dengan oversampling. Hasil terbaik diperoleh pada skenario kedua dengan penerapan oversampling menggunakan metode RandomOverSampling, yang menghasilkan akurasi sebesar 88.66% dan F1-score sebesar 89.19%. Untuk penelitian selanjutnya, ada kemungkinan untuk mengembangkan metode Jaringan Syaraf Tiruan dengan Particle Swarm Optimization ini dengan menggabungkannya dengan metode klasifikasi lainnya. Selain itu, penelitian dapat diperluas dengan menguji lebih banyak skenario untuk mencapai hasil performa yang lebih baik. Disarankan juga untuk melakukan eksplorasi lebih lanjut tentang metode Jaringan Syaraf Tiruan dan Particle Swarm Optimization untuk meningkatkan pemahaman tentang potensi dan aplikasi mereka.

Daftar Pustaka

- [1] K. Jual Beli Akun Twitter and A. Mahasiswa Program Studi Sarjana Ilmu Hukum Bagian Hukum Perdata dan Ekonomi, "KEABSAHAN JUAL BELI AKUN TWITTER DITINJAU DARI KITAB UNDANG-UNDANG HUKUM PERDATA," *Jurnal Ilmiah Kutei*, vol. 21, no. 1. pp. 15–27, Jul. 20, 2022. Accessed: Jul. 19, 2023. [Online]. Available: <https://ejournal.unib.ac.id/jkutei/article/view/22847>
- [2] F. Zahria Emeraldien, R. Jefri Sunarsono, R. Alit, J. Raya Rungkut Madya, G. Anyar, and J. Timur, "TWITTER SEBAGAI PLATFORM KOMUNIKASI POLITIK DI INDONESIA," 2019. [Online]. Available: www.statisticbrain.com
- [3] J. Nayak, H. Swapnarekha, B. Naik, G. Dhiman, and S. Vimal, "25 Years of Particle Swarm Optimization: Flourishing Voyage of Two Decades," *Archives of Computational Methods in Engineering*, vol. 30, no. 3, pp. 1663–1725, Apr. 2023, doi: 10.1007/S11831-022-09849-X/METRICS.
- [4] H. Amarullah Purwaatmaja Ash-Shidiq EFSA and K. Muslim Lhaksmana, "Prediksi Retweet Menggunakan Fitur Berbasis Pengguna dan Fitur Berbasis Konten dengan Metode Klasifikasi ANN."
- [5] M. S. Syah Zannuar and K. Muslim Lhaksmana, "Prediksi Retweet Berdasarkan Feature User-Based Menggunakan Metode Klasifikasi Random Forest," 2021.
- [6] I. Puspita Dewi and K. Muslim Lhaksmana, "Prediksi Retweet Menggunakan Metode Bernoulli dan Gaussian Naive Bayes di Media Sosial Twitter Dengan Topik Vaksinasi Covid-19," 2021.
- [7] M. Rizqi Akbar, "Prediksi Retweet Berdasarkan User-Based dan Content-Based Menggunakan Metode Ensemble Stacking," 2023.
- [8] K. Muslim Lhaksmana, "Prediksi retweet berdasarkan feature user-based menggunakan metode klasifikasi Support Vector Machine," 2021.
- [9] D. Duei Putri, G. F. Nama, and W. E. Sulistiono, "Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (DPR) Pada Twitter Menggunakan Metode Naive Bayes Classifier," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 10, no. 1, Jan. 2022, doi: 10.23960/jitet.v10i1.2262.
- [10] B. Hauriza and I. M. Wirawan, "Prediksi Tingkat Inflasi Bulanan Indonesia Menggunakan Metode Jaringan Saraf Tiruan," *Jurnal Teknologi dan Informasi (JATI)*, vol. 11, no. 2, p. 123, 2021, doi: 10.34010/jati.v11i2.
- [11] J. P. Haumahu, K. Kunci, : Backpropagation, J. S. Tiruan, N. Balok, and P. Pola, "Implementasi Jaringan Syaraf Tiruan Untuk Pengenalan Pola Notasi Balok Menggunakan Metode Backpropagation," 2019. [Online]. Available: <http://ejournal.stmik-budidarma.ac.id/index.php/jurikom|Page255>
- [12] F. Dristyan, "Seminar Nasional Royal (SENAR) 2018 ISSN 2622-9986 (cetak) STMIK Royal-AMIK Royal, hlm. 185-190 ISSN 2622-6510 (online) Kisaran, Asahan," 2018.
- [13] T. Wardati Khusniyah, "Prediksi Nilai Tukar Petani Menggunakan Jaringan Syaraf Tiruan Backpropagation," *Scientific Journal of Informatics*, vol. 3, no. 1, 2016, [Online]. Available: <http://journal.unnes.ac.id/nju/index.php/sji>
- [14] S. Kawatra and A. Vaishnava, "Detection of Choroidalneovascularization (Cnv) in Retina Oct Images Using Vgg16 and Densenet Cnn," 2021, doi: 10.21203/rs.3.rs-360517/v1.
- [15] Y. Eka Achyani STMIK Nusa Mandiri Jakarta, "Penerapan Metode Particle Swarm Optimization Pada Optimasi Prediksi Pemasaran Langsung," *JURNAL INFORMATIKA*, vol. 5, no. 1, pp. 1–11, 2018.
- [16] Y. P. R. S. Ardiana Rosita, "Implementasi Algoritma Particle Swarm untuk Menyelesaikan Sistem Persamaan Nonlinear," 2012, Accessed: Jul. 21, 2023. [Online]. Available: 10.12962/j23373539.v1i1.636
- [17] N. Febri Istighfarin, R. Anisa Rahmastati, and H. Nugroho, "PENERAPAN METODE PARTICLE SWARM OPTIMIZATION (PSO) DAN GENETIC ALGORITHM (GA) PADA SISTEM OPTIMASI VISIBLE LIGHT COMMUNICATION (VLC) UNTUK MENENTUKAN POSISI ROBOT," *Jurnal SIMETRIS*, vol. 11, no. 1, 2020.
- [18] D. Normawati and S. A. Prayogi, "Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter," 2021.
- [19] F. Matematika, I. Pengetahuan, A. Dan, T. Informasi, and J. Pendidikan, "Ika Menarianti," *Jurnal Ilmiah Teknosains*, vol. 1, no. 1, 2015.