

Integrasi Strategi *Pre-processing* Data untuk Optimalisasi Akurasi Algoritma *Backpropagation*

Widodo Saputra¹, Saifullah², Eka Irawan³, Anjar Wanto^{4*}

^{1,2,3,4*} STIKOM Tunas Bangsa, Pematangsiantar, Indonesia

¹widodo@amiktunasbangsa.ac.id, ²saifullah@amiktunasbangsa.ac.id, ³eka.irawan@amiktunasbangsa.ac.id,

^{4*}anjarwanto@amiktunasbangsa.ac.id

*) anjarwanto@amiktunasbangsa.ac.id

Abstrak—Algoritma Backpropagation merupakan salah satu metode jaringan syaraf tiruan yang banyak digunakan dalam proses klasifikasi dan prediksi karena kemampuannya dalam mengenali pola data secara akurat. Namun, performa algoritma ini sangat dipengaruhi oleh kualitas data masukan. Data yang tidak terstruktur, memiliki perbedaan skala, mengandung *missing value*, maupun fitur yang kurang relevan dapat menurunkan tingkat akurasi model. Penelitian ini bertujuan untuk menganalisis pengaruh integrasi strategi *pre-processing* data terhadap optimalisasi akurasi algoritma Backpropagation. Dataset yang digunakan berasal dari Badan Pusat Statistik (BPS) berupa data Tingkat Pengangguran Terbuka (TPT) penduduk usia 15 tahun ke atas di Provinsi Sumatera Utara tahun 2019–2024. Tahapan *pre-processing* yang diterapkan meliputi data *cleaning*, normalisasi, penanganan *missing value*, dan reduksi fitur. Metode penelitian dilakukan dengan membandingkan hasil pengujian model menggunakan *pre-processing* standar dan *pre-processing* parsial pada beberapa arsitektur jaringan. Hasil penelitian menunjukkan bahwa penerapan strategi *pre-processing* mampu meningkatkan performa model Backpropagation. Nilai akurasi tertinggi diperoleh pada arsitektur 3-56-1 dengan peningkatan dari 80,00% menjadi 85,88%. Selain meningkatkan akurasi, proses pelatihan model juga menjadi lebih stabil dan konvergensi error dicapai lebih cepat. Dengan demikian, integrasi strategi *pre-processing* data terbukti efektif dalam mengoptimalkan akurasi algoritma Backpropagation pada permasalahan prediksi berbasis data numerik.

Kata Kunci: *Backpropagation*, *pre-processing* data, jaringan syaraf tiruan, akurasi, prediksi

Abstract—*Backpropagation* is one of the artificial neural network algorithms widely used in classification and prediction processes due to its ability to recognize data patterns accurately. However, the performance of this algorithm is highly influenced by the quality of the input data. Unstructured data, differences in data scales, missing values, and irrelevant features can reduce the model's accuracy. This study aims to analyze the effect of integrating data *pre-processing* strategies to optimize the accuracy of the Backpropagation algorithm. The dataset used in this research was obtained from the Badan Pusat Statistik (BPS) in the form of Open Unemployment Rate data for the population aged 15 years and above in North Sumatra Province from 2019 to 2024. The applied *pre-processing* stages included data *cleaning*, normalization, missing value handling, and feature reduction. The research method was conducted by comparing the model testing results using standard *pre-processing* and partial *pre-processing* on several network architectures. The results showed that the implementation of *pre-processing* strategies was able to improve the performance of the Backpropagation model. The highest accuracy value was obtained in the 3-56-1 architecture with an increase from 80.00% to 85.88%. In addition to improving accuracy, the model training process became more stable and the error convergence was achieved faster. Therefore, the integration of data *pre-processing* strategies has proven to be effective in optimizing the accuracy of the Backpropagation algorithm for numerical data-based prediction problems

Keywords: *Backpropagation*, data *pre-processing*, artificial neural network, accuracy, prediction.

1. PENDAHULUAN

Perkembangan teknologi informasi dan komputasi cerdas telah mendorong pemanfaatan *Machine Learning* (ML) di berbagai sektor, mulai dari prediksi finansial, diagnosis medis, hingga pengenalan pola [1]. Jaringan syaraf tiruan (JST) dengan algoritma *backpropagation* adalah algoritma pembelajaran yang paling umum digunakan dalam untuk mengoptimalkan bobot jaringan berdasarkan perhitungan error [2]. Meskipun algoritma klasik seperti jaringan saraf tiruan telah digunakan untuk prediksi, algoritma tersebut seringkali kesulitan dalam menangani data yang kompleks dan mengidentifikasi pola dalam kumpulan data yang besar [3][4]. JST merupakan bagian dari *Artificial Intelligent* dimana *Artificial Intelligence* merupakan cabang ilmu komputer yang mempelajari cara menciptakan mesin atau komputer agar mampu menjalankan tugas layaknya manusia, bahkan dengan hasil yang dapat melebihi kemampuan manusia dalam beberapa aspek. Perkembangan *artificial intelligence* mampu memberikan terobosan-terobosan yang sangat inovatif mengikuti kondisi terkini[5]. Bersamaan dengan itu, Kecerdasan Buatan (AI) mewakili domain lain yang berkembang pesat, memperoleh stabilitas dari waktu ke waktu

[6]. Salah satu metode *problem solving* dari JST adalah menyelesaikan kasus prediksi dengan beberapa metode seperti *perceptron*, *backpropagation* dan *fuzzy* [7].

Backpropagation merupakan metode populer dalam bidang *machine learning* karena kemampuannya dalam mengenali pola dan melakukan prediksi yang akurat [8][9]. *Backpropagation* bekerja dengan cara meminimalkan error menggunakan algoritma *gradient descent* untuk memperbaiki bobot jaringan secara iteratif. Namun, performa *backpropagation* sangat bergantung pada kualitas data yang digunakan [10]. Data yang tidak terstruktur, memiliki skala berbeda, mengandung *outlier*, atau terdapat nilai yang hilang, dapat menyebabkan proses pelatihan menjadi tidak optimal, lambat dalam mencapai konvergensi, serta menghasilkan akurasi prediksi yang rendah. Fenomena "*garbage in, garbage out*" menjadi tantangan utama, di mana data mentah (*raw data*) yang diperoleh dari dunia nyata umumnya bersifat kotor, tidak konsisten, mengandung penciran (*outliers*), nilai yang hilang (*missing values*), serta perbedaan skala antar-fitur yang masif. Oleh karena itu, diperlukan penerapan strategi *data preprocessing* yang tepat untuk meningkatkan efektivitas pelatihan dan hasil dari algoritma *backpropagation* [11]. *Data preprocessing* mencakup berbagai teknik yang bertujuan untuk mempersiapkan data agar lebih sesuai digunakan dalam proses pelatihan jaringan syaraf tiruan [12]. Beberapa strategi yang umum digunakan antara lain normalisasi untuk menyamakan rentang nilai variabel, standarisasi agar distribusi data lebih seimbang, reduksi dimensi untuk mengurangi kompleksitas perhitungan, serta penanganan *outlier* agar tidak mendistorsi pola utama yang ingin dipelajari [13]. Untuk menangani permasalahan tersebut, proses pra-pemrosesan data (*data pre-processing*) merupakan tahapan yang sangat penting sebelum data dimasukkan ke dalam jaringan saraf [14]. Berbagai metode pra-pemrosesan konvensional, seperti normalisasi data menggunakan *Min-Max Scaling* maupun *Z-score Standardization*, telah terbukti dapat mempercepat proses komputasi dengan cara menyamakan rentang nilai pada data masukan [15].

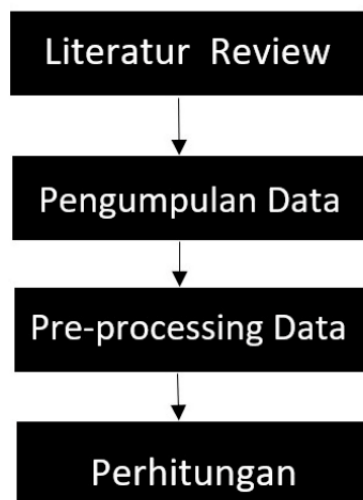
Dalam penerapannya, *backpropagation* dapat menyelesaikan beberapa kasus seperti prediksi data. Dalam penelitian ini data yang diuji adalah data yang berasal dari Badan Pusat Statistik (BPS) berupa data Tingkat Pengangguran Terbuka (TPT) penduduk usia 15 tahun ke atas di Provinsi Sumatera Utara tahun 2019–2024. Tingkat Pengangguran Terbuka (TPT) adalah salah satu indikator makroekonomi penting yang menggambarkan kondisi perekonomian suatu negara sekaligus tingkat kesejahteraan sosial masyarakat [16].

Pengangguran adalah keadaan ketika seseorang yang termasuk dalam angkatan kerja dan memiliki keinginan untuk bekerja belum memperoleh pekerjaan. Kondisi ini tidak hanya berdampak pada individu, tetapi juga memengaruhi perekonomian daerah secara menyeluruh. [17]. Pengangguran merupakan salah satu permasalahan besar yang selalu diperhatikan bangsa Indonesia dari tahun ke tahun selain kemiskinan. Pengangguran saat ini merupakan masalah signifikan yang harus segera diselesaikan. Seseorang yang sedang mencari pekerjaan, belum memiliki pekerjaan sama sekali, mencari pekerjaan yang sesuai, atau hanya bekerja selama dua hari dalam seminggu dapat dikategorikan sebagai pengangguran. Pengangguran umumnya terjadi karena jumlah pencari kerja tidak seimbang dengan jumlah lapangan pekerjaan yang tersedia [18]. Tingkat pengangguran yang tinggi menyebabkan penurunan daya beli masyarakat, peningkatan kemiskinan, dan menurunnya produktivitas regional [17]. Pengangguran dapat timbul akibat ketidakseimbangan dalam pasar tenaga kerja, di mana jumlah tenaga kerja yang tersedia melebihi permintaan tenaga kerja yang ada [19]. Jika masalah pengangguran terus terjadi, kondisi tersebut dapat memberikan dampak buruk terhadap pertumbuhan ekonomi. Meningkatnya tingkat pengangguran biasanya menyebabkan perlambatan pertumbuhan ekonomi akibat menurunnya jumlah output atau produksi yang dihasilkan [19]. Dengan demikian, kemampuan dalam memprediksi TPT secara tepat menjadi alat yang sangat penting bagi pemerintah dan para pengambil kebijakan untuk menyusun strategi ketenagakerjaan, menentukan alokasi anggaran, serta merancang program perlindungan sosial yang bersifat preventif.

Melalui data yang digunakan, data akan dibagi menjadi 2 (dua) yaitu data latih dan data uji dimana data tahun 2019 sampai 2021 menjadi data latih pada target tahun 2022. Data tahun 2021 sampai 2024 menjadi data uji pada target tahun 2024. Untuk tahapan *pre-processing* yang diterapkan meliputi data *cleaning*, normalisasi, penanganan *missing value*, dan reduksi fitur dengan mengubah pola *pre-processing* standard dengan modifikasi *pre-processing*. Penerapan strategi *preprocessing* yang sistematis, proses pelatihan *backpropagation* dapat menjadi lebih stabil, konvergensi error lebih cepat, dan hasil prediksi lebih akurat. Penelitian ini bertujuan untuk menganalisis pengaruh integrasi beberapa strategi *pre-processing data* terhadap peningkatan akurasi algoritma *Backpropagation*. Secara keseluruhan, penelitian ini dapat memberikan kontribusi dalam pengembangan metode *machine learning*, khususnya pada penerapan *backpropagation*. Hasil yang diperoleh dapat menjadi acuan bagi peneliti maupun praktisi dalam memilih strategi *data preprocessing* yang sesuai dengan karakteristik data yang digunakan. Dengan demikian, model jaringan syaraf tiruan tidak hanya memiliki kemampuan prediksi yang lebih tinggi, tetapi juga lebih andal untuk digunakan dalam berbagai aplikasi, seperti pengenalan pola, klasifikasi, maupun prediksi.

2. METODE PENELITIAN

Metode penelitian ini mencakup serangkaian tahapan yang harus dilaksanakan dalam membangun sistem prediksi, dimulai dari kajian literatur, pengumpulan data, pra-pemrosesan data, hingga proses perhitungan seperti ditunjukkan pada gambar 1.



Gambar 1. Tahap Penelitian

2.1 Pengumpulan Data

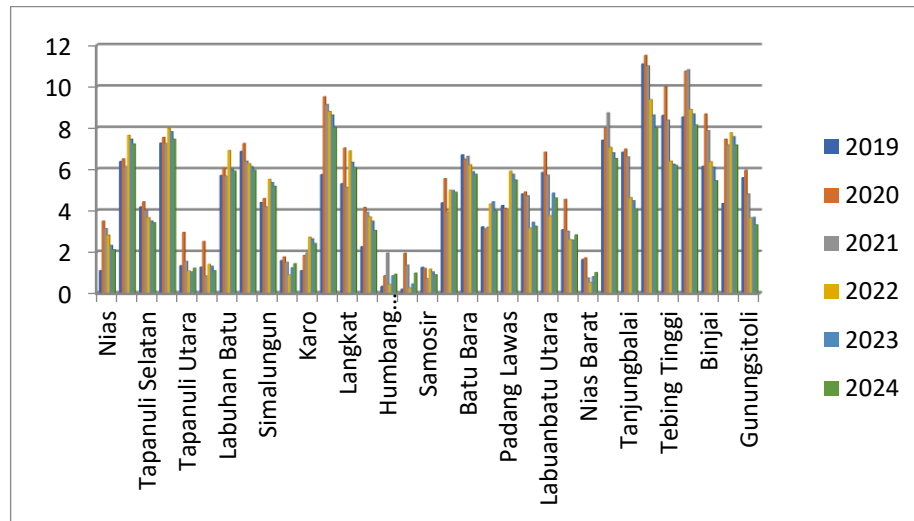
Dataset yang digunakan dalam penelitian ini bersumber dari Badan Pusat Statistik (BPS) yaitu data Tingkat Pengangguran Terbuka (TPT) Penduduk Umur 15 Tahun Keatas Menurut Kab/Kota pada Provinsi Sumatera Utara. Dataset dapat dilihat pada tabel 1.

Tabel 1. Data TPT Penduduk Umur 15 Tahun Keatas Provinsi Sumatera Utara

Kota/Kab	2019	2020	2021	2022	2023	2024
Nias	1,09	3,49	3,12	2,81	2,31	2,1
Mandailing Natal	6,37	6,5	6,12	7,64	7,45	7,22
Tapanuli Selatan	4,17	4,42	4	3,65	3,49	3,41
Tapanuli Tengah	7,26	7,54	7,24	7,97	7,81	7,45
Tapanuli Utara	1,33	2,94	1,54	1,07	1,03	1,21
Toba	1,26	2,5	0,83	1,39	1,3	1,09
Labuhan Batu	5,7	6,05	5,66	6,9	5,99	5,9
Asahan	6,86	7,24	6,39	6,26	6,12	5,94
Simalungun	4,39	4,58	4,17	5,51	5,35	5,17
Dairi	1,58	1,75	1,49	0,88	1,23	1,43
Karo	1,09	1,83	1,95	2,71	2,63	2,4
Deli Serdang	5,74	9,5	9,13	8,79	8,62	8,02
Langkat	5,3	7,02	5,12	6,88	6,33	6,08
Nias Selatan	2,25	4,15	3,91	3,69	3,48	3,03
Humbang Hasundutan	0,33	0,84	1,94	0,42	0,84	0,92
Pakpak Bharat	0,19	1,93	1,36	0,26	0,45	0,97
Samosir	1,25	1,2	0,7	1,16	1,03	0,89
Serdang Bedagai	4,37	5,54	3,93	4,98	4,97	4,88

Kota/Kab	2019	2020	2021	2022	2023	2024
Batu Bara	6,69	6,48	6,62	6,21	5,88	5,75
Padang Lawas Utara	3,21	3,11	3,19	4,31	4,42	3,99
Padang Lawas	4,24	4,11	4,07	5,9	5,75	5,47
Labuhanbatu Selatan	4,8	4,9	4,71	3,15	3,43	3,24
Labuanbatu Utara	5,84	6,82	5,71	3,75	4,84	4,6
Nias Utara	3,07	4,54	3	2,59	2,57	2,82
Nias Barat	1,63	1,71	0,74	0,53	0,8	1
Sibolga	7,4	8	8,72	7,05	6,79	6,52
Tanjungbalai	6,82	6,97	6,59	4,62	4,47	4,08
Pematangsiantar	11,09	11,5	11	9,36	8,62	8
Tebing Tinggi	8,6	9,98	8,37	6,39	6,24	6,18
Medan	8,53	10,74	10,81	8,89	8,67	8,13
Binjai	6,14	8,67	7,86	6,36	6,1	5,44
Padangsidempuan	4,34	7,45	7,18	7,76	7,57	7,17
Gunungsitoli	5,59	5,94	4,8	3,65	3,67	3,3

Dari tabel 1 dapat dilihat bahwa terdapat 33 Kabupaten/Kota yang ada di Sumatera Utara. Menurut BPS pengangguran terbuka dapat terdiri dari mereka yang tak punya pekerjaan dan mencari pekerjaan[20]. TPT adalah perbandingan total yang tidak bekerja terhadap total angkatan kerja [21]. Grafik perbandingan TPT umur 15 tahun keatas dapat dilihat pada gambar 2.



Gambar 2. Grafik Perbandingan TPT

2.2 Pre-processing Data

Tahap ini mencakup beberapa langkah dasar untuk memastikan data dalam format yang optimal untuk pelatihan dalam meningkatkan nilai akurasi sehingga dapat memaksimalkan hasil prediksi. Tahap *pre-processing* dilakukan untuk memastikan bahwa data yang akan digunakan telah sesuai dengan kebutuhan model pembelajaran mesin [22][23].

2.3 Perhitungan Backpropagation

Backpropagation merupakan metode pelatihan yang dilakukan secara terstruktur untuk meningkatkan kinerja jaringan saraf tiruan yang memiliki lebih dari satu lapisan [24]. Seperti yang sudah dibahas sebelumnya metode *Backpropagation* ini memiliki 4 tahap yaitu tahap *initialization*, tahap *activation*, tahap *bobot training*, dan tahap *iteration* [25]. Untuk lebih jelasnya dapat dilihat pada penjelasan berikut ini:

Iterasi 1

Langkah 0

Inisiasi semua bobot dengan bilangan acak.

- a. Inisiasi bobot dan bias pada *hidden layer*.

Pada lapisan input layer setelah dilakukan penginputan data untuk terhubung dengan lapisan *hidden layer* dan dari *hidden layer* ke *output layer* diperlukan nilai bobot dan nilai bias. Nilai bobot dan bias ini bisa saja berubah ketika melakukan proses training di aplikasi *Matlab* yang digunakan.

- b. Bobot pada *hidden layer* ke *output layer*

Penginputan nilai bobot dari lapisan *hidden layer* menuju *output layer* (W_{jk}) dimana nilai ini didapatkan secara acak. Setiap komputer ketika menentukan bobot nilainya bisa berubah-ubah.

Fase I : *Feedforward*

Langkah 3

Tiap unit input menerima sinyal dan meneruskan ke *hidden layer*.

Langkah 4

Menghitung nilai *output* dari unit tersembunyi ataupun *hidden layer* (Z_j).

Pada tahapan ini dilakukan perhitungan dari setiap bobot pada masing-masing layer sesuai dengan fungsi aktivasi *sigmoid* pada metode *Backpropagation*. Adapun langkah-langkahnya sebagai berikut:

1. Setiap unit pada lapisan *input layer* ($X_i, i = 1, 2, 3, \dots, p$) menerima sinyal X_i dan meruskan ke semua pada lapisan *hidden layer*.
2. Setiap unit pada lapisan *hidden layer* ($X_i, i = 1, 2, 3, \dots, p$) menjumlahkan nilai bobot dengan nilai input dengan persamaan sebagai berikut:

$$z_{netj} = v_{j0} + \sum_{i=1}^n x_i v_{ji} \quad (1)$$

Menghitung semua output pada lapisan *hidden layer* menggunakan fungsi aktivasi, perhitungannya sebagai berikut:

$$Z_1 = \frac{1}{1 + e^{-(Z_{netj})}} \quad (2)$$

Langkah 5

Menghitung hasil pada unit *ouput* yk

$$y_{netk} = w_{0k} + \sum_{i=1}^p z_i w_{jk} \quad (3)$$

Menghitung *output* dengan fungsi aktivasi :

$$Y_1 = \frac{1}{1 + e^{-(Z_{inj})}} \quad (4)$$

Fase II : *Backforward*

Langkah 6

Menghitung nilai faktor δ pada unit *output* berdasarkan nilai kesalahan disetiap unit keluaran.

$$\delta_k = (t_k - y_k) y_k (1 - y_k)$$

Menghitung koreksi bobot dengan persamaan sebagai berikut:

$$\Delta w_{jk} = \alpha \delta_k Z_j, \text{ dimana untuk learning ratenya } 0.1$$

Langkah 7

Menghitung faktor δ pada unit tersembunyi berdasarkan kesalahan pada unit tersembunyi Z_j dengan persamaan:

$$\delta_{netj} = \sum_{k=1}^n \delta_k w_{jk} \quad (5)$$

Untuk menghitung informasi kesalahan/error, maka nilai tersebut dikalikan dengan turunan dari fungsi aktivasi dengan persamaan:

$$\delta_j = \delta_{netj} z_j Z_j (1 - z_j)$$

Kemudian hitung koreksi bobot yang nantinya akan digunakan untuk memperbaiki nilai V_{ji} .

$$\Delta v_{ji} = \alpha \delta_j x_i$$

Fase III *Upgrade Bobot*

Langkah 8

Perubahan bobot yang menuju *hidden layer*, dengan persamaan:

$$V_{ji} \text{ (baru)} = v_{ji} \text{ (lama)} + \Delta v_{ji}$$

Ubah bobot menuju *output layer*, dengan persamaan:

$$w_{kj} \text{ (baru)} = w_{kj} \text{ (lama)} + \Delta w_{kj}$$

Setelah didapatkan bobot akhir maka bisa selanjutnya akan di bandingkan dengan goal yang telah dibuat, jika belum memenuhi maka dilakukan iterasi selanjutnya. Iterasi berhenti jika telah mencapai nilai goal yang

ditentukan atau sudah mendekati. Karena pada iterasi ke 1 belum tercapai dengan goal yang diharapkan maka dilakukan proses iterasi ke 2 dan selanjutnya.

3. HASIL DAN PEMBAHASAN

Penelitian ini bertujuan untuk menganalisis pengaruh integrasi beberapa strategi *pre-processing data* terhadap peningkatan akurasi algoritma Backpropagation. Strategi *pre-processing* yang diterapkan meliputi **pembersihan data (data cleaning)**, **normalisasi**, **penanganan data hilang**, dan **reduksi fitur**. Eksperimen dilakukan dengan membandingkan performa model Backpropagation sebelum dan sesudah dilakukan *pre-processing data*.

3.1 Hasil

a. Hasil Pengujian Dengan *Pre-processing* Standar

Pada tahap awal, algoritma Backpropagation dilatih menggunakan data mentah tanpa dilakukan *pre-processing*. Hasil pengujian dilakukan dengan 6 (enam) kali percobaan dan menunjukkan hasil akurasi yang relatif rendah. Hal ini disebabkan oleh adanya skala data yang tidak seragam, nilai kosong (*missing value*), serta fitur yang kurang relevan sehingga memengaruhi proses pembelajaran jaringan saraf tiruan.

Hasil perhitungan pengujian menggunakan Pre-processing Standar dapat dilihat dari tabel 2.

Tabel 2. Hasil Pengujian Dengan Pre-processing Standar

No	Metode	Arsitektur	Epoch	Times	Accuracy (%)
1	Backpropagation Pre-processing Standar	3-5-1	4245	12	44,12
2		3-27-1	1426	5	64,18
3		3-43-1	924	6	72,35
4		3-33-1	2075	5	62,35
5		3-56-1	2360	7	80,00
6		3-63-1	960	9	78,35

b. Hasil Pengujian dengan *Pre-processing* Parsial

Pengujian selanjutnya dilakukan dengan menerapkan *pre-processing* secara parsial, seperti modifikasi normalisasi data. Hasilnya menunjukkan adanya peningkatan akurasi dibandingkan data mentah. Normalisasi membantu mempercepat proses konvergensi dan mengurangi kesalahan pada proses pembaruan bobot.

Hasil perhitungan pengujian menggunakan Pre-processing Parsial dapat dilihat dari tabel 3 dibawah.

Tabel 3. Hasil Pengujian Dengan *Pre-processing* Parsial

No	Metode	Arsitektur	Epoch	Times	Accuracy (%)
1	Modifikasi Pengolahan Data Backpropagation	3-5-1	4287	11	50
2		3-27-1	1441	4	55,29
3		3-43-1	928	2	72,35
4		3-33-1	2072	5	65,29
5		3-56-1	2369	6	85,88
6		3-63-1	929	2	79,29

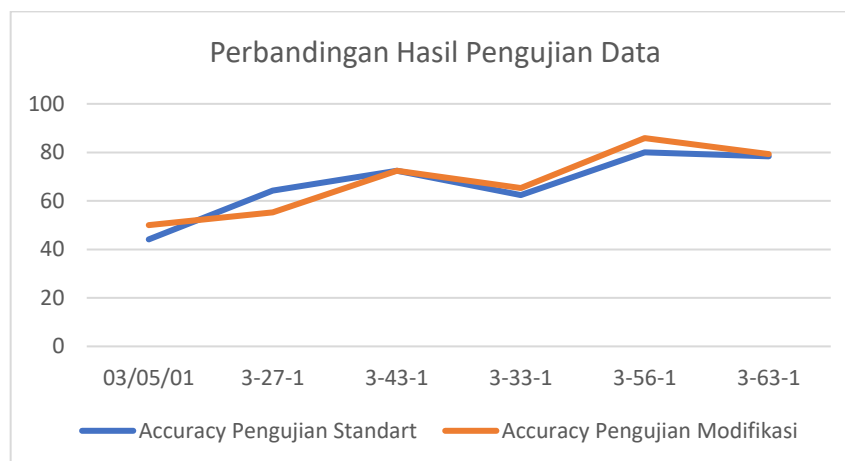
c. Perbandingan Hasil Pengujian Data

Berikut hasil perbandingan pengujian data yang dihasilkan dari pengujian standart dan pengujian yang telah dimodifikasi dengan *Pre-processing* Parsial. Hasil perbandingan dapat dilihat dari table 4 dibawah.

Tabel 4. Perbandingan Hasil Pengujian Data

No	Arsitektur	Accuracy Pengujian Standart	Accuracy Pengujian Modifikasi	Status
1	3-5-1	44,12	50	Meningkat
2	3-27-1	64,18	55,29	Menurun
3	3-43-1	72,35	72,35	Tetap
4	3-33-1	62,35	65,29	Meningkat
5	3-56-1	80,00	85,88	Meningkat
6	3-63-1	78,35	79,29	Meningkat

Dari tabel 4 dapat dilihat perbandingan nilai akurasi dari hasil pengujian data antara pengujian standar dengan pengujian modifikasi. Pada pola 3-5-1 nilai akurasi meningkat dari 44,12% menjadi 50%, pada pola 3-20-1 terjadi penurunan nilai akurasi yaitu dari 41,18% menjadi 35,29%, pada pola 3-37-1 nilai akurasi yang didapat adalah tetap. Pada pola selanjutnya yaitu 3-19-1, 3-26-1, 3-35-1 seluruhnya mengalami peningkatan. Berikut grafik perbandingan hasil pengujian data.



Gambar 3. Perbandingan Hasil Pengujian Data

Grafik Perbandingan Hasil Pengujian Data menunjukkan perbandingan tingkat akurasi antara Accuracy Pengujian Standart dan Accuracy Pengujian Modifikasi pada beberapa skenario pengujian data. Garis biru merepresentasikan hasil pengujian standar, sedangkan garis oranye menunjukkan hasil pengujian modifikasi.

Berdasarkan grafik, pada pengujian pertama dengan label 03/05/01, metode modifikasi memperoleh nilai akurasi sekitar 50%, lebih tinggi dibandingkan metode standar yang berada di kisaran 45%. Pada pengujian kedua (3-27-1), akurasi metode standar meningkat cukup signifikan hingga sekitar 64%, sedangkan metode modifikasi berada pada nilai sekitar 55%.

Selanjutnya, pada pengujian ketiga (3-43-1), kedua metode menunjukkan hasil yang hampir sama dengan akurasi sekitar 72%. Pada pengujian keempat (3-33-1), kedua metode mengalami sedikit penurunan akurasi, namun metode modifikasi tetap lebih unggul dibandingkan metode standar.

Peningkatan akurasi tertinggi terjadi pada pengujian kelima (3-56-1), di mana metode modifikasi mencapai sekitar 86%, sedangkan metode standar mencapai sekitar 80%. Pada pengujian terakhir (3-63-1), kedua metode mengalami sedikit penurunan, namun metode modifikasi masih memberikan hasil yang lebih tinggi dibandingkan metode standar.

Secara keseluruhan, grafik menunjukkan bahwa metode pengujian modifikasi cenderung menghasilkan tingkat akurasi yang lebih baik dibandingkan metode standar pada sebagian besar skenario pengujian. Hal ini mengindikasikan bahwa metode modifikasi mampu meningkatkan performa sistem dalam proses pengujian data.

3.2 Pembahasan

Berdasarkan serangkaian pengujian dan analisis yang telah dilakukan pada proses penelitian ini, dapat disimpulkan bahwa strategi *pre-processing* data memiliki pengaruh yang sangat signifikan dalam meningkatkan performa dan optimalisasi akurasi algoritma Backpropagation. Tahapan *pre-processing* bukan hanya berfungsi sebagai proses

awal pengolahan data, tetapi juga menjadi faktor penting yang menentukan kualitas hasil pembelajaran pada jaringan saraf tiruan (*Artificial Neural Network*). Data mentah yang digunakan secara langsung tanpa melalui proses *pre-processing* umumnya masih mengandung berbagai permasalahan, seperti perbedaan skala nilai antar atribut, data yang hilang (*missing value*), data tidak relevan, serta noise yang dapat mengganggu proses pelatihan model. Kondisi tersebut menyebabkan jaringan saraf mengalami kesulitan dalam mengenali pola dan hubungan antar data secara optimal, sehingga tingkat akurasi prediksi atau klasifikasi menjadi kurang maksimal.

Hasil pengujian menunjukkan bahwa proses normalisasi data mampu memberikan dampak positif terhadap proses pembelajaran algoritma Backpropagation. Normalisasi membantu menyamakan rentang nilai dari setiap fitur atau atribut sehingga tidak terdapat perbedaan nilai yang terlalu jauh antar variabel. Dengan rentang data yang lebih seimbang, proses pembaruan bobot (*weight update*) pada jaringan saraf dapat berjalan lebih stabil dan efektif. Selain itu, normalisasi juga membantu mempercepat proses konvergensi selama pelatihan sehingga model dapat mencapai tingkat kesalahan (*error*) yang lebih kecil dalam waktu yang lebih efisien.

Penanganan *missing value* juga terbukti memberikan kontribusi penting terhadap peningkatan kualitas data. Data yang memiliki nilai kosong atau tidak lengkap dapat menyebabkan gangguan dalam proses pembelajaran dan menghasilkan prediksi yang kurang akurat. Melalui teknik penanganan *missing value*, data menjadi lebih konsisten dan siap digunakan dalam proses pelatihan. Hal ini mampu mengurangi noise pada dataset sehingga risiko terjadinya *overfitting* maupun kesalahan klasifikasi dapat diminimalkan. Dengan data yang lebih bersih dan terstruktur, jaringan saraf dapat mempelajari pola data dengan lebih baik.

Selain itu, proses reduksi fitur (*feature reduction*) memberikan dampak positif dalam menyederhanakan kompleksitas data. Penghapusan atribut yang tidak relevan atau memiliki kontribusi kecil terhadap hasil prediksi membantu algoritma fokus pada fitur-fitur yang benar-benar penting. Dengan jumlah fitur yang lebih optimal, proses komputasi menjadi lebih ringan dan efisien, serta mampu meningkatkan kemampuan model dalam mengenali pola utama pada dataset. Reduksi fitur juga membantu mengurangi kemungkinan redundansi data yang dapat memengaruhi performa jaringan saraf.

Secara keseluruhan, hasil penelitian ini menunjukkan bahwa penerapan strategi *pre-processing* data secara terintegrasi merupakan langkah yang sangat efektif dalam meningkatkan performa algoritma Backpropagation, khususnya pada permasalahan klasifikasi dan prediksi berbasis data numerik. Kombinasi antara normalisasi data, penanganan *missing value*, dan reduksi fitur terbukti mampu menghasilkan model yang lebih stabil, akurat, dan efisien. Oleh karena itu, proses *pre-processing* data sebaiknya menjadi tahapan utama yang tidak dapat diabaikan dalam pengembangan sistem berbasis jaringan saraf tiruan agar hasil prediksi yang diperoleh menjadi lebih optimal dan dapat diandalkan.

4. KESIMPULAN

Berdasarkan hasil penelitian dan pembahasan yang telah dilakukan, dapat disimpulkan bahwa integrasi strategi *pre-processing data* berpengaruh signifikan terhadap peningkatan akurasi algoritma Backpropagation. Penerapan tahapan *pre-processing* seperti pembersihan data, normalisasi, penanganan *missing value*, dan reduksi fitur mampu meningkatkan kualitas data input sehingga proses pembelajaran jaringan saraf tiruan berjalan lebih optimal.

Hasil pengujian menunjukkan bahwa penggunaan data yang telah melalui *pre-processing* terintegrasi menghasilkan nilai akurasi yang lebih tinggi dan tingkat kesalahan (*loss*) yang lebih rendah dibandingkan dengan penggunaan data mentah maupun *pre-processing* parsial. Selain itu, proses pelatihan model menjadi lebih stabil dan konvergensi dicapai dalam jumlah iterasi yang lebih sedikit.

Dengan demikian, penelitian ini membuktikan bahwa strategi *pre-processing data* merupakan faktor penting dalam meningkatkan performa algoritma Backpropagation, khususnya pada permasalahan klasifikasi dan prediksi berbasis data numerik.

UCAPAN TERIMA KASIH

Penulis menyampaikan ucapan terima kasih yang tulus dan penghargaan setinggi-tingginya kepada Pengurus Yayasan STIKOM Tunas Bangsa Pematangsiantar atas dukungan finansial dan fasilitas pendanaan penelitian yang diberikan melalui program Hibah Penelitian Internal. Ucapan terima kasih juga disampaikan kepada **STIKOM Tunas Bangsa** yang telah memberikan dukungan akademik, pendampingan, serta sarana dan prasarana penelitian.

REFERENSI

- [1] K. Razzaq and M. Shah, "Machine Learning and Deep Learning Paradigms: From Techniques to Practical Applications and Research Frontiers," *Computers*, vol. 14, no. 3, p. 93, 2025, doi: 10.3390/computers14030093.



- [2] P. Prihandoko and P. Alkhairi, "Optimasi JST Backpropagation dengan Adaptive Learning Rate Dalam Memprediksi Hasil Panen Padi," *Jurasik (Jurnal Ris. Sist. Inf. dan Tek. Inform.*, vol. 10, no. 1, p. 441, 2025, doi: 10.30645/jurasik.v10i1.887.
- [3] S. Solikhun and M. R. Siregar, "Analyzing Perceptron Algorithm for Global Gold Price Prediction using Quantum Computing Approach," *Kinet. Game Technol. Inf. Syst. Comput. Network, Comput. Electron. Control*, vol. 4, no. 1, 2025, doi: 10.22219/kinetik.v10i1.2024.
- [4] M. Yanto, S. Sanjaya, Yulasmi, D. Guswandi, and S. Arlis, "Implementation multiple linear regresion in neural network predict gold price," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 22, no. 3, pp. 1635–1642, 2021, doi: 10.11591/ijeecs.v22.i3.pp1635-1642.
- [5] R. Pakpahan, "Analisa Pengaruh Implementasi Artificial," *J. Inf. Syst. Informatics Comput.*, vol. 5, no. 2, pp. 506–513, 2021, doi: 10.52362/jisicom.v5i2.616.
- [6] T. Baidawi *et al.*, "Comparison Of Adaline And Hebbian Algortihms On Pattern," vol. 10, no. 1, pp. 23–31, 2024, doi: 10.33480/jitk.v10i1.4941.COMPARISON.
- [7] A. P. Windarto, S. Defit, and A. Wanto, "Optimalisasi Parameter dengan Cross Validation dan Neural Back-propagation Pada Model Prediksi Pertumbuhan Industri Mikro dan Kecil," *J. Sist. Inf. Bisnis*, vol. 11, no. 1, pp. 34–42, 2021, doi: 10.21456/vol11iss1pp34-42.
- [8] E. Sofiati, "Optimasi Prediksi Pencetakan E-Ktp: Penerapan Revolusioner Jaringan Syaraf Tiruan Dengan Metode Backpropagation," *JIKO (Jurnal Inform. dan Komputer)*, vol. 9, no. 1, p. 68, 2025, doi: 10.26798/jiko.v9i1.1420.
- [9] S. Gustina, "Aplikasi Machine Learning untuk Mendeteksi Kematangan Tomat menggunakan Metode Backpropagation," *J. Engine Energi, Manufaktur, dan Mater.*, vol. 8, no. 1, pp. 81–88, 2024, doi: 10.30588/jeemm.v8i1.1815.
- [10] I. Permana and F. N. S. Salisah, "The Effect of Data Normalization on the Performance of the Classification Results of the Backpropagation Algorithm," *Indones. J. Inform. Res. Softw. Eng.*, vol. 2, no. 1, pp. 67–72, 2022.
- [11] F. Marwati and R. Fauzi, "Prediksi Penyakit Diabetes Melitus Menggunakan Jaringan Syaraf Tiruan Dengan Metode Backpropagation," *Jitu J. Inform. Utama Hal*, vol. 2, no. 1, pp. 26–34, 2024.
- [12] D. Irmansyah, "Jurnal JISIILKOM (Jurnal Inovasi Sistem Informasi & Ilmu Komputer) Prediksi Penyakit Paru-Paru Menggunakan Jaringan Syaraf Tiruan Dengan Metode Backpropagation," *J. JISIILKOM*, vol. 3, no. 1, pp. 3025–4868, 2025.
- [13] D. Kurniasari, M. N. A. Rafdiono, and W. Warsono, "Analisis Struktur Terbaik Neural Network dengan Algoritma Backpropagation dalam Memprediksi Indeks Kandungan Sulfida (SO₂) di Ibu Kota Jakarta," *J. Sist. dan Teknol. Inf.*, vol. 12, no. 2, p. 321, 2024, doi: 10.26418/justin.v12i2.76166.
- [14] R. N. Azizah and D. Avianto, "TIN : Terapan Informatika Nusantara Implementasi Convolutional Neural Network Untuk Pengenalan Tulisan Tangan Akasara Sunda Ngalangéna TIN : Terapan Informatika Nusantara," vol. 6, no. 7, pp. 1037–1049, 2025, doi: 10.47065/tin.v6i7.8703.
- [15] P. P. Alloserung, A. Erna, M. Bagussahrir, and S. Alam, "Analisis Performa Normalisasi Data untuk Klasifikasi K-Nearest Neighbor pada Dataset Penyakit," *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 9, no. 3, pp. 178–191, 2024, doi: 10.14421/jiska.2024.9.3.178-191.
- [16] N. Ulfiyah, D. D. Sari, and S. Ningsih, "ANALISIS TINGKAT PENGANGGURAN TERBUKA (TPT) PROVINSI JAWA," vol. 3, pp. 376–384, 2026.
- [17] J. Halif, D. Wahiddin, I. Sanjaya, and S. Faisal, "Model Regresi Linear Berganda untuk Prediksi Tingkat Pengangguran di Provinsi Jawa Barat," *J. Algoritma.*, vol. 22, no. 1, pp. 324–335, 2025, doi: 10.33364/algoritma/v.22-1.2312.
- [18] N. Amalya and S. Solikhun, "Polak-Ribiere Conjugate Gradient Algorithm in Predicting the Percentage of Open Unemployment in North Sumatra Province," *J. Tek. Inform.*, vol. 5, no. 1, pp. 31–38, 2024, doi: 10.52436/1.jutif.2024.5.1.1047.
- [19] D. A. N. Upah, M. Terhadap, P. Di, J. Barat, Q. N. Ramadhani, and C. F. Ananda, "Jdess 03.02.2024," vol. 3, no. 2, pp. 347–361, 2024.
- [20] D. K. P. Wee, G. M. V. Kawung, and I. Masloman, "Analisis Faktor-Faktor yang Mempengaruhi Tingkat Pengangguran Terbuka dan Kemiskinan di Kabupaten Minahasa," *J. Berk. Ilm. Efisiensi*, vol. 24, no. 2, pp. 133–145, 2024.
- [21] L. Hakim, Hailuddin, and E. Agustiani, "Pengaruh Tingkat Pengangguran Terbuka Terhadap Jumlah Kemiskinan Di Provinsi Ntb Pada Tahun 2010-2021," *J. Econ. Bus.*, vol. 10, no. 1, pp. 21–32, 2024, doi: 10.29303/ekonobis.v10i1.158.
- [22] A. Gunawan, A. Hermawan, and D. Avianto, "BULLETIN OF COMPUTER SCIENCE RESEARCH Analisis Pengaruh Preprocessing Data dan Hyperparameter Tuning pada Backpropagation Neural Network dalam Klasifikasi Stroke," vol. 6, no. 2, pp. 682–692, 2026, doi: 10.47065/bulletincsr.v6i2.956.
- [23] M. Resa Arif Yudianto, P. Sukmasetya, R. Abul Hasani, and D. Sasongko, "Pengaruh Data Preprocessing



- terhadap Imbalanced Dataset pada Klasifikasi Citra Sampah menggunakan Algoritma Convolutional Neural Network,” *Build. Informatics, Technol. Sci.*, vol. 4, no. 3, pp. 1367–1375, 2022, doi: 10.47065/bits.v4i3.2575.
- [24] M. I. Ulumando, “Model Prediksi Prestasi Mahasiswa ITAF Kupang Menggunakan Backpropagation Berbasis Data Perilaku Belajar,” *J. Nas. Komputasi dan Teknol. Inf.*, vol. 8, no. 4, pp. 1770–1781, 2025.
- [25] R. T. Untari and M. Devegi, “Penerapan Algoritma Backpropagation untuk Memprediksi Jumlah Permintaan Buku dan Alat Tulis,” *JURTEII J. Teknol. Inf.*, vol. 2, no. 1, pp. 1–7, 2022, doi: 10.22202/jurteii.2022.6670.