

MODEL REGRESI LOGISTIK DAN JARINGAN SYARAF TIRUAN UNTUK KLASIFIKASI MAHASISWA BERPOTENSI DROP OUT

Armansyah¹⁾, Suhardi²⁾

^{1,2}Program Studi Ilmu Komputer

^{1,2}Fakultas Sains dan Teknologi

^{1,2}Universitas Islam Negeri Sumatera Utara

E-mail: armansyah@uinsu.ac.id¹⁾, suhardi@uinsu.ac.id²⁾

Abstract

Every university faces students who leave without notice, including those who fail to complete their studies and are declared as dropouts (DO). An initial step in addressing student dropout issues can be undertaken using classification techniques. This study aims to classify dropout students using logistic regression, which is compared with the Artificial Neural Network (ANN) method to categorize data into five classifications: Active, Graduated, Potential to Graduate, Potential DO, and DO. The dataset consists of academic records of undergraduate students from the Computer Science program, obtained from PUSTIPADA at UIN Sumatera Utara. The data includes entry year, study duration, semester GPA, cumulative GPA, credits per semester, total credits, and tuition fees. A total of 1,337 student records were divided into 80% training and 20% testing sets. The logistic regression model achieved an accuracy of 93% on the test data, while the ANN model performed better with an accuracy of 96%. This indicates that ANN is more effective in capturing complex and variable patterns in student data. The findings of this study contribute to academic institutions and educational policymakers, particularly in the Computer Science program, by providing insights for decision-making and developing intervention programs to prevent potential dropouts among students with similar characteristics to those in the dataset.

Keywords: Drop out, Undergraduate students, Logistic Regression, Artificial Neural Network.

Intisari

Setiap perguruan tinggi memiliki mahasiswa yang keluar tanpa keterangan, termasuk mahasiswa yang gagal hingga akhir masa studinya dan dinyatakan *drop out* (DO). Langkah awal penyelesaian permasalahan mahasiswa DO dapat dilakukan melalui teknik klasifikasi. Penelitian ini bertujuan untuk mengklasifikasi mahasiswa DO menggunakan metode regresi logistik yang dikomparasi dengan metode Jaringan Syaraf Tiruan (JST) untuk membagi data dalam 5 klasifikasi: Aktif, Lulus, Potensi Lulus, Potensi DO, dan DO. Data yang digunakan adalah data akademik mahasiswa Program Studi Ilmu Komputer yang diperoleh dari PUSTIPADA UIN Sumatera Utara yang berupa: tahun masuk, lama studi, IPS, IPK, SKS per semester, total SKS, dan UKT. Sebanyak 1337 data mahasiswa yang dibagi dalam 80% pelatihan, dan 20% pengujian. Hasil pengujian model regresi logistik terhadap data uji memperoleh akurasi 93%, sementara model JST lebih unggul dengan akurasi 96%. Hal ini menunjukkan bahwa JST lebih efektif dalam menangkap pola kompleks dan variatif pada data mahasiswa. Hasil penelitian ini dapat memberikan kontribusi kepada lembaga akademik dan pengambil kebijakan pendidikan khususnya Program Studi Ilmu Komputer untuk membuat keputusan atau program pencegahan potensi *drop out* terhadap mahasiswa yang memiliki karakteristik serupa dengan data yang digunakan dalam penelitian ini.

Kata Kunci: *Drop out*, Mahasiswa, Regresi Logistik, Jaringan Syaraf Tiruan.

1. PENDAHULUAN

Machine learning merupakan bagian dari kecerdasan buatan yang banyak digunakan untuk memecahkan berbagai masalah, termasuk dalam bidang pendidikan. *Machine learning* dapat didefinisikan sebagai aplikasi komputer

dan algoritma matematika yang belajar dari data untuk menghasilkan klasifikasi di masa yang akan datang[1]. Beberapa metode yang sering digunakan dalam *machine learning* untuk klasifikasi adalah regresi logistik dan Jaringan Syaraf Tiruan (JST).

Regressi logistik adalah model matematis yang digunakan untuk memberikan klasifikasi berdasarkan hubungan antara variabel bebas dan variabel kategorikal[2], [3]. Regressi logistik multinomial memungkinkan analisis hubungan antara variabel prediktor dengan variabel respon yang memiliki lebih dari dua kategori[3]. Di sisi lain, JST adalah sistem komputasi yang meniru cara kerja jaringan saraf biologis manusia, dengan karakteristik pengolahan paralel, pengolahan data dalam jumlah besar, serta toleransi terhadap kesalahan[4]. JST terdiri dari lapisan masuk, tersembunyi, dan keluar, yang bekerja melalui propagasi maju dan mundur untuk melatih model hingga mencapai hasil optimal[5].

JST telah terbukti mampu meningkatkan efektivitas klasifikasi data dengan menerapkan teknik pelatihan yang lebih fleksibel dan mampu menangkap pola yang lebih kompleks dibandingkan metode tradisional[6]. Sebagai contoh, studi oleh Tapas Si et al.[7] menunjukkan bahwa JST yang dioptimalkan dengan metaheuristik dapat meningkatkan akurasi klasifikasi data medis, sehingga mempercepat proses analisis dan prediksi dalam berbagai bidang. Dalam konteks klasifikasi akademik, JST juga dapat digunakan secara lebih universal dengan struktur data tertentu, serta memungkinkan pengambilan keputusan yang lebih akurat dalam mendeteksi pola risiko mahasiswa *Drop Out* (DO) sejak dini.

Penerapan kedua metode ini juga telah dilakukan dalam berbagai penelitian pendidikan, misalnya untuk memprediksi nilai siswa berdasarkan data nilai akademiknya[8]. Dengan kemampuan klasifikasi dari regressi logistik dan JST, perguruan tinggi dapat menganalisis data mahasiswa untuk mendeteksi potensi DO lebih awal dan merancang langkah preventif [9], [10].

Setiap perguruan tinggi menghadapi masalah DO yang sering terjadi dan membutuhkan solusi efektif untuk mengatasinya. DO mengacu pada situasi ketika mahasiswa gagal menyelesaikan studi akibat berbagai faktor, seperti masalah finansial, nilai IPK yang rendah, atau perilaku yang tidak sesuai[11]. Masalah ini tidak hanya berdampak pada mahasiswa itu sendiri, tetapi juga pada reputasi dan efisiensi institusi pendidikan. Pada penelitian terkait potensi DO sebelumnya, para peneliti umumnya hanya

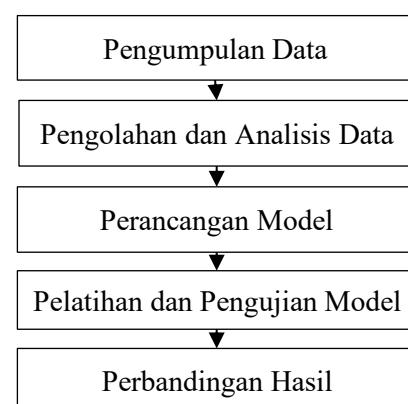
berfokus pada klasifikasi biner untuk mewakili potensi DO dan tidak DO, sehingga belum ada pendekatan yang mencakup klasifikasi yang lebih rinci (multi kelas).

Metode regresi logistik dan JST telah diterapkan dalam berbagai penelitian klasifikasi dan prediksi. Misalnya, Fauziyah menggunakan regresi logistik untuk menganalisis risiko penyakit jantung coroner[12]. Fahmuddin menerapkannya untuk klasifikasi berita dengan akurasi 78,3%[13]. Sementara itu, Komariah memanfaatkan JST-Backpropagation untuk memprediksi hasil produksi kelapa sawit dengan akurasi 88% dan *Mean Squared Error* (MSE) sebesar 0,11249[14].

Penelitian ini bertujuan untuk membandingkan performa regresi logistik dan JST dalam mengklasifikasi mahasiswa ke dalam lima kategori akademik: Aktif, Lulus, Potensi Lulus, Potensi DO, dan DO, berdasarkan data akademik mahasiswa. Dengan pendekatan multikelas, penelitian ini memungkinkan analisis yang lebih mendalam untuk mendeteksi potensi DO sejak dini[15]. Hasil penelitian diharapkan memberikan kontribusi dalam merancang program pencegahan DO yang berbasis data, guna meningkatkan keberhasilan pendidikan tinggi.

2. METODOLOGI

Penelitian ini mengikuti kerangka metodologi yang terdiri dari lima langkah utama berikut gambar 1:



Gambar 1. Kerangka Penelitian

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari PUSTIPADA (Pusat Teknologi Informasi dan Pangkalan Data) Universitas Islam Negeri Sumatera Utara (Sugiyono dalam [16]).

Pemilihan sumber data ini dilakukan karena PUSTIPADA memiliki rekam jejak akademik mahasiswa yang terstruktur dan terdokumentasi dengan baik, mencakup variabel yang relevan dengan penelitian ini. Data yang digunakan mencakup mahasiswa Program Studi Ilmu Komputer angkatan 2015 hingga 2021 dengan total 1337 mahasiswa, sehingga memberikan cakupan yang luas dalam menganalisis pola akademik dan faktor yang berpengaruh terhadap potensi DO. Pemilihan dataset ini juga didasarkan pada kebutuhan untuk mengeksplorasi keterkaitan antara variabel akademik dengan kemungkinan mahasiswa mengalami DO, mendukung pengembangan model klasifikasi berbasis JST. Variabel yang digunakan dalam penelitian ini meliputi Nomor Induk Mahasiswa (NIM), tahun masuk, lama studi dalam semester, Indeks Prestasi Semester (IPS), Indeks Prestasi Kumulatif (IPK), jumlah Satuan Kredit Semester (SKS) per semester, total SKS terakhir, biaya kuliah Uang Kuliah Tunggal (UKT), serta label kategori potensi DO yang terdiri dari lima kelas, yaitu aktif(0), lulus(1), potensial lulus(2), potensial DO(3), dan DO(4). Sebelum digunakan dalam pemodelan, data yang diperoleh dari PUSTIPADA mengalami proses pembersihan untuk memastikan kualitas dan konsistensinya. Proses ini dilakukan secara manual menggunakan *Microsoft Excel*, mencakup penghapusan data duplikat, penanganan nilai yang hilang (*missing values*), serta pemeriksaan konsistensi format data. Tahapan ini bertujuan untuk meningkatkan kualitas data sehingga model klasifikasi dapat memberikan hasil prediksi yang lebih akurat dan dapat diandalkan.

Data yang telah dikumpulkan melalui proses pembersihan secara manual menggunakan *Microsoft excel*, selanjutnya dinormalisasikan secara *min-max* [17] dengan persamaan :

$$x_{new} = \frac{x_{old} - x_{min}}{x_{max} - x_{min}} \quad (1)$$

Dimana, x_{new} adalah nilai fitur yang telah dinormalisasi, x_{old} adalah nilai asli, x_{min} adalah nilai minimum fitur, dan x_{max} adalah nilai maksimum fitur. Proses ini memastikan

semua fitur berada dalam rentang yang seragam untuk meningkatkan stabilitas analisis.

Selanjutnya dengan data hasil normalisasi tersebut, koefisien bobot dihitung dengan memaksimalkan fungsi *log-likelihood* menggunakan persamaan :

$$L(w) = \sum_{i=1}^m \sum_{j=1}^n y_{ij} \log P(y = j|x_i) \quad (2)$$

Dalam perhitungan *log-likelihood* regresi logistik multinomial, m adalah jumlah sampel, dan n adalah jumlah kelas (dalam kasus ini 5 kelas). Probabilitas kelas $P(y=j|x_i)$, yaitu probabilitas data ke- i termasuk kelas j , dihitung sebagai :

$$P(y = j|x_i) = \frac{e^{x_i^T w_j}}{\sum_{k=1}^n e^{x_i^T w_k}} \quad (3)$$

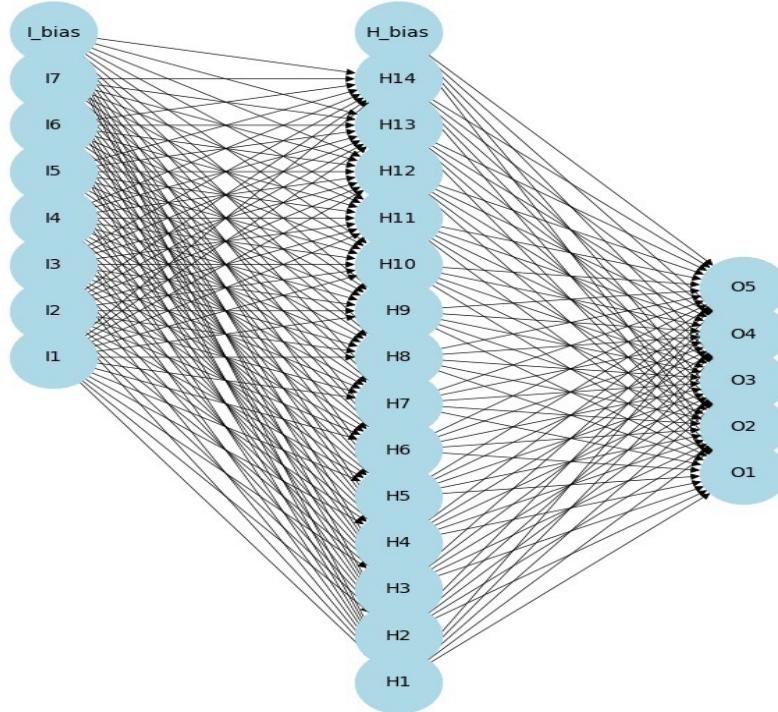
di mana x_i adalah fitur data yang telah dinormalisasi, dan w_j adalah koefisien bobot untuk kelas j . Tahap akhir prediksi dilakukan dengan menentukan kelas j yang memiliki probabilitas maksimum berdasarkan data fitur x . Rumus prediksi kelas adalah:

$$Y_{prediksi} = \operatorname{argmax}_j P(y = j|x) \quad (4)$$

Kelas dengan nilai $P(y=j|x)$ tertinggi dipilih sebagai hasil prediksi.

Adapun untuk analisis dengan JST, tahapan yang dilakukan adalah sebagai berikut:

1. Persiapan Data: Data yang telah dinormalisasi digunakan sebagai input untuk model JST, dengan fitur-fitur seperti tahun masuk, lama studi, IPS, IPK, SKS, total SKS, dan biaya kuliah.
2. Membangun Model: dengan arsitektur 7-14-5, yaitu 7 input neuron, 14 neuron di lapisan tersembunyi dengan fungsi aktivasi ReLU, dan 5 *output* neuron menggunakan fungsi aktivasi *Softmax*. Gambar 2 menunjukkan arsitektur model JST untuk penelitian ini.



Gambar 2. Arsitektur Model JST (7-14-5)

3. Pelatihan Model: Model dilatih menggunakan *learning rate* 0.009 selama 1000 epoch, dengan pembaruan bobot dan bias melalui algoritma backpropagation yang mencakup fase *feedforward*, perhitungan error, dan pembaruan *parameter* bobot dan bias.

Fase Feedforward:

Dimana fase ini menghitung output dari jaringan berdasarkan input yang diberikan dan bobot yang ada ditambah dengan bias masing-masing lapisan.

Pada lapisan tersembunyi dan lapisan *output*, net (z_j) dihitung dengan rumus:

$$z_j = \sum_{n=1}^N x_i w_{ij} + b_j \quad (5)$$

$$a_j = \max(0, z_j) \quad (6)$$

Nilai input x_i dikalikan dengan bobot w_{ij} dan ditambah bias b_j , hasilnya diaktivasikan dengan fungsi ReLU (a_j = output lapisan tersembunyi). Sedangkan pada lapisan akhir, *output* (Y_j) diaktivasi dengan *softmax* menggunakan persamaan berikut :

$$Y_j = \frac{e^{z_j}}{\sum_{k=1}^m e^k} \quad (7)$$

Perhitungan Error:

Fase ini menggunakan fungsi *cross-entropy loss* untuk menghitung *error* antara *output* prediksi dan label yang sebenarnya. *Error (Loss)* dihitung dengan persamaan :

$$L = - \sum_{i=1}^N \sum_{j=1}^K Y_{ij} \log(Y'_{ij}) \quad (8)$$

Dalam perhitungan fungsi *error* diatas, N adalah jumlah sampel dan K adalah jumlah kelas. Y_{ij} merupakan nilai aktual untuk kelas j pada sampel i , sementara Y'_{ij} adalah prediksi untuk kelas j pada sampel i .

Pembaruan Bobot dan Bias:

Fase pembaruan bobot dan bias dilakukan dengan menggunakan *Gradien Descent* untuk meminimalkan *error*, yakni dengan rumus :

$$w_{ij}^{new} = w_{ij}^{old} - \eta * \frac{\partial L}{\partial w_{ij}} \quad (9)$$

$$b_j^{new} = b_j^{old} - \eta * \frac{\partial L}{\partial b_j} \quad (10)$$

η adalah *learning rate* yang mengatur perubahan bobot dan bias, sementara $\frac{\partial L}{\partial w_{ij}}$ dan $\frac{\partial L}{\partial b_j}$ adalah turunan fungsi *Error* terhadap bobot dan bias yang dihitung melalui backpropagation.

Setelah model dilatih, prediksi dilakukan pada data uji, dan hasilnya dievaluasi dengan akurasi dan *confusion matrix*. Hasil dari JST dibandingkan dengan regresi logistik untuk menentukan model yang lebih efektif dalam memprediksi potensi drop out mahasiswa.

3. HASIL DAN PEMBAHASAN

Bagian ini memaparkan hasil evaluasi kinerja model regresi logistik dan JST dalam memprediksi mahasiswa yang berpotensi DO. Gambar 3 menunjukkan dataset yang digunakan dalam penelitian ini. Dataset selanjutnya dinormalisasikan dengan Min-Max, yang

seperti pada gambar 4. Hasil normalisasi data digunakan untuk pelatihan model. Setelah melalukan pelatihan data, pada model Regresi Logistik diperoleh nilai koefisien regresi seperti yang terdaftar pada gambar 5.

Hasil evaluasi dari model Regresi Logistik dan JST, baik pada tahap pelatihan maupun pengujian, dibahas secara rinci setelah paragraf ini, termasuk akurasi, *precision*, *recall*, dan f1-score, untuk memberikan gambaran menyeluruh mengenai performa masing-masing model dalam mengidentifikasi mahasiswa dengan risiko DO.

	nim	tahun_masuk	lama_studi	ips	ipk	skssemester	total_sks	biaya_kuliah	label
0	71153001	2015	14	0.00	2.77	0	166	2400000	1
1	71153005	2015	14	0.00	2.75	0	174	2400000	1
2	71153006	2015	14	0.00	2.88	0	164	2400000	1
3	71153009	2015	14	0.00	2.08	6	198	2400000	2
4	71153014	2015	14	0.00	2.73	0	154	2400000	4
...
1332	701213246	2021	2	2.89	0.74	18	38	2400000	3
1333	701213247	2021	2	3.52	1.80	21	41	2926000	0
1334	701213248	2021	2	3.05	2.78	21	41	1463000	0
1335	701213250	2021	2	0.00	0.00	0	0	1463000	4
1336	701213251	2021	2	3.24	1.61	21	41	2194000	0

[1337 rows x 9 columns]

Gambar 3. Dataset Model

Normalisasi:

```
[[ 1.35573744 -1.35573744 -0.18121672 ... 0.91846656 -1.19644419
-0.99961015]
[ 0.76874541 -0.76874541  0.61063241 ...  1.12850844 -0.37529479
-0.29003019]
[ -0.40523865  0.40523865 -1.12652498 ... -1.28697315 -1.62657007
-1.78206191]
...
[ 1.35573744 -1.35573744  1.06487145 ...  0.91846656 -1.19644419
-0.46153655]
[ 0.76874541 -0.76874541  0.6842928 ...  0.91846656 -0.4534995
-0.88404578]
[ 1.35573744 -1.35573744  0.91755068 ...  0.91846656 -1.19644419
-0.46153655]]
```

Gambar 4. Hasil Normalisasi Data

Kelas	Fitur	Koefisien	Kelas	Fitur	Koefisien	Kelas	Fitur	Koefisien
Kelas 0	Tahun Masuk	2.13	Kelas 1	Tahun Masuk	-1.60	Kelas 2	Tahun Masuk	1.06
	Lama Studi	-2.13		Lama Studi	1.60		Lama Studi	-1.06
	IPS	1.66		IPS	-1.18		IPS	0.44
	IPK	-0.87		IPK	2.21		IPK	1.03
	SKS Semester	2.33		SKS Semester	-0.56		SKS Semester	0.28
	Total SKS	-1.79		Total SKS	1.08		Total SKS	4.49
	Biaya Kuliah	-0.68		Biaya Kuliah	0.35		Biaya Kuliah	0.03
Kelas	Fitur	Koefisien	Kelas	Fitur	Koefisien			
Kelas 3	Tahun Masuk	-0.28	Kelas 4	Tahun Masuk	-1.31			
	Lama Studi	0.28		Lama Studi	1.31			
	IPS	-0.16		IPS	-0.77			
	IPK	-1.03		IPK	-1.34			
	SKS Semester	-0.58		SKS Semester	-1.47			
	Total SKS	-0.58		Total SKS	-3.21			
	Biaya Kuliah	-0.04		Biaya Kuliah	0.35			

Gambar 5. Koefisien Regresi Model

Pada tahap pelatihan, model regresi logistik mencapai akurasi sebesar 92%, yang menunjukkan performa yang cukup baik dalam mengklasifikasikan mahasiswa berdasarkan status akademiknya. Kelas mayoritas seperti Aktif (kelas 0) dan Potensi Lulus (kelas 2) memiliki performa yang sangat baik dengan nilai *precision*, *recall*, dan f1-score mendekati atau mencapai 1.00. Hal ini mengindikasikan bahwa model sangat efektif dalam mengenali mahasiswa yang aktif atau berpotensi lulus.

Namun, pada kelas minoritas seperti Potensi DO (kelas 3) dan Lulus (kelas 1), performa model menurun secara signifikan, terutama pada *recall*. Nilai *recall* untuk kelas 3 hanya sebesar 7%, menunjukkan bahwa model kesulitan mengenali mahasiswa dengan potensi DO.

Hal yang sama terjadi pada tahap pengujian seperti yang ditunjukkan pada gambar 6., di mana akurasi model stabil di angka 93%, tetapi masalah yang sama pada kelas minoritas tetap terlihat. Matriks konfusi menunjukkan bahwa mahasiswa dalam kelas Potensi DO sering salah diklasifikasikan sebagai Potensi Lulus (kelas 2).

Distribusi *precision* dan *recall* yang tidak merata menunjukkan adanya ketidakseimbangan data antar kelas, yang berdampak pada kemampuan model dalam memprediksi kelas minoritas. Dalam hal ini diperlukan strategi tambahan seperti metode penyeimbangan data atau pemberian bobot kelas untuk meningkatkan performa model pada kelas-kelas yang ada.

Akurasi : 0.93

Klasifikasi Pengujian:		precision	recall	f1-score	support
	0	1.00	0.99	0.99	89
	1	0.80	0.62	0.70	13
	2	0.93	0.99	0.96	117
	3	0.40	0.22	0.29	9
	4	0.90	0.90	0.90	40
		accuracy		0.93	268
		macro avg	0.81	0.74	0.77
		weighted avg	0.92	0.93	0.93

Matriks Konfusi:	
[[88 0 0 0 1]	
[0 8 4 0 1]	
[0 0 116 1 0]	
[0 0 1 4 2 2]	
[0 1 1 2 36]]	

Gambar 6. Hasil dan Evaluasi Pengujian Model Regresi Logistik

Adapun hasil pelatihan model JST menunjukkan performa yang sangat baik dengan akurasi pelatihan sebesar 96%. Hal ini menunjukkan bahwa model dapat mengklasifikasikan status akademik mahasiswa dengan tingkat keberhasilan yang tinggi. Seperti halnya pada model regresi logistik, kelas Aktif (kelas 0) dan Potensi Lulus (kelas 2) tetap mendominasi dengan hasil *precision*, *recall*, dan f1-score yang hampir sempurna, terutama pada tahap pelatihan. *Precision* dan *recall* masing-masing mencapai nilai 1.00 untuk kelas 0 dan 0.99 untuk kelas 2, yang mencerminkan kemampuan model dalam mengenali mahasiswa aktif dan yang berpotensi lulus dengan sangat baik.

Pada kelas Potensi DO (kelas 3), model JST memiliki performa yang sedikit lebih baik

dibandingkan dengan regresi logistik. Meskipun nilai *recall* masih terbilang rendah, yaitu 61% pada pelatihan dan 67% pada pengujian, terdapat peningkatan yang signifikan dalam kemampuan model untuk mengidentifikasi mahasiswa dengan potensi DO. Precision untuk kelas ini juga menunjukkan nilai 0.85 pada pelatihan dan 0.75 pada pengujian, lebih baik dibandingkan dengan model regresi logistik yang lebih kesulitan.

Pada kelas Lulus (kelas 1), meskipun precision tetap berada pada angka 0.73 pada pelatihan dan 0.75 pada pengujian, terdapat sedikit penurunan *recall* pada pengujian (69%), yang dapat diartikan bahwa meskipun model cukup baik dalam mendeteksi mahasiswa yang lulus, masih ada sejumlah data yang salah diklasifikasikan.

Matriks konfusi pada kedua fase (pelatihan dan pengujian) memperlihatkan bahwa kelas Aktif (kelas 0) dan Potensi Lulus (kelas 2) dideteksi dengan sangat baik, sementara kelas Potensi DO (kelas 3) dan Lulus (kelas 1) masih memiliki tingkat kesalahan yang perlu diperbaiki.

Terkait dengan pelatihan, nilai *error* pada setiap *epoch* menunjukkan penurunan yang stabil dan cenderung mencapai titik konvergen pada epoch ke-300 dengan nilai *error* tetap bernilai 0.1353 seperti yang ditunjukkan pada gambar 8. Kondisi ini mengindikasikan bahwa model telah mencapai titik stabilitas dalam proses pelatihan.

Secara keseluruhan, model JST menunjukkan performa yang lebih unggul dibandingkan dengan regresi logistik dalam hal akurasi dan kemampuan dalam mengenali kelas tertentu, meskipun beberapa kelas minoritas masih memerlukan perhatian lebih lanjut dalam hal penyeimbangan dan peningkatan *recall*. Strategi seperti penyeimbangan data dengan metode *weight class* atau regularisasi tambahan bisa dipertimbangkan untuk lebih meningkatkan performa pada kelas-kelas dengan jumlah data terbatas.

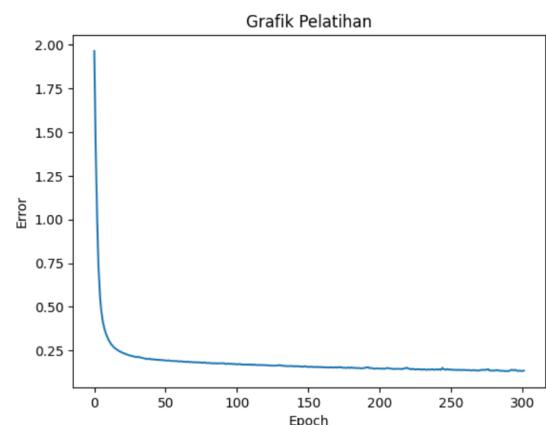
Gambar 7 menyajikan hasil dan evaluasi pengujian model JST, yang menunjukkan bahwa meskipun akurasi keseluruhan cukup tinggi, terdapat ruang untuk perbaikan pada kelas-kelas yang memiliki distribusi data yang tidak seimbang, terutama dalam meningkatkan *recall* pada kelas dengan jumlah data lebih sedikit.

Laporan Klasifikasi Pengujian:				
	precision	recall	f1-score	support
0	1.00	0.99	0.99	89
1	0.75	0.69	0.72	13
2	0.96	0.98	0.97	117
3	0.75	0.67	0.71	9
4	0.97	0.97	0.97	40
accuracy			0.96	268
macro avg	0.89	0.86	0.87	268
weighted avg	0.96	0.96	0.96	268

Confusion Matrix Pengujian:

[[88 0 0 1 0]]
[0 9 3 0 1]
[0 1 115 1 0]
[0 1 2 6 0]
[0 1 0 0 39]]

Gambar 7. Hasil dan Evaluasi Pengujian Model Jaringan Saraf Tiruan



Gambar 8. Grafik Pelatihan Model Jaringan Saraf Tiruan

Berdasarkan hasil evaluasi yang telah dijelaskan sebelumnya, perbandingan antara model regresi logistik dan JST menunjukkan perbedaan dalam beberapa metrik performa. Tabel 1 memperlihatkan bahwa JST memiliki akurasi yang lebih tinggi baik pada tahap pelatihan maupun pengujian dibandingkan dengan regresi logistik. Secara khusus, JST mencapai akurasi 0.96 pada kedua tahap, sementara regresi logistik memiliki akurasi 0.92 pada pelatihan dan 0.93 pada pengujian.

Tabel 1. Perbandingan Akurasi dan Precision

Metode	Tahap	Precision					
		Akurasi	Kelas 0	Kelas 1	Kelas 2	Kelas 3	Kelas 4
Regresi Logistik	Pelatihan	0.92	0.99	0.74	0.91	0.67	0.88
	Pengujian	0.93	1.00	0.75	0.91	1.00	0.88
JST	Pelatihan	0.96	1.00	0.73	0.95	0.85	0.99
	Pengujian	0.96	1.00	0.75	0.96	0.75	0.97

Tabel 2. Perbandingan Recall dan F1-Score

Metode	Tahap	Recall						F1-Score			
		Kelas 0	Kelas 1	Kelas 2	Kelas 3	Kelas 4	Kelas 0	Kelas 1	Kelas 2	Kelas 3	Kelas 4
Regresi Logistik	Pelatihan	1.00	0.64	1.00	0.07	0.94	1.00	0.69	0.95	0.13	0.91
	Pengujian	0.99	0.46	0.99	0.11	0.95	0.99	0.57	0.95	0.20	0.92
JST	Pelatihan	1.00	0.80	0.99	0.61	0.95	1.00	0.76	0.97	0.71	0.97
	Pengujian	0.99	0.69	0.98	0.67	0.97	0.99	0.72	0.97	0.71	0.97

Tabel 2 menunjukkan hasil yang lebih rinci tentang kemampuan kedua model dalam mengenali kelas-kelas tertentu. Pada metrik *precision*, *recall*, dan *f1-score*, JST menunjukkan kinerja yang lebih stabil dan lebih tinggi dibandingkan dengan regresi logistik. Terutama pada kelas-kelas dengan data minoritas, JST dapat menghasilkan prediksi yang lebih akurat. Meskipun ada beberapa kelas yang masih memerlukan perhatian lebih, khususnya pada kelas 3 untuk *recall*, secara keseluruhan JST menawarkan performa yang lebih seimbang dan unggul.

4. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian ini, dapat disimpulkan bahwa model JST menunjukkan performa yang lebih baik dibandingkan dengan model Regresi Logistik dalam hal akurasi, precision, recall, dan f1-score. Hasil evaluasi menunjukkan bahwa JST mampu mencapai akurasi 96% pada tahap pengujian, lebih tinggi dibandingkan dengan Regresi Logistik yang mencapai 93%. JST juga lebih efektif dalam menangani kelas-kelas minoritas, meskipun masih ada beberapa kelas dengan nilai recall yang perlu diperbaiki. Sementara itu, model Regresi Logistik, meskipun lebih sederhana, tetap memberikan hasil yang baik, namun memiliki kesulitan dalam mengklasifikasikan beberapa kelas tertentu, terutama Kelas 1 dan Kelas 3. Penelitian ini menunjukkan bahwa

untuk kasus klasifikasi dengan banyak kelas dan ketidakseimbangan data, model JST lebih unggul.

Namun, pengembangan lebih lanjut masih diperlukan pada kedua model untuk meningkatkan kemampuan keduanya dalam mengenali kelas minoritas dengan lebih akurat. Penggunaan metode penyeimbangan data seperti oversampling (misalnya SMOTE atau ADASYN), *undersampling*, atau penerapan class weights pada fungsi loss dapat menjadi solusi efektif. Selain itu, pendekatan hybrid seperti kombinasi oversampling dan undersampling, augmented data synthesis, atau *Generative Adversarial Networks* (GANs) untuk menciptakan data tambahan juga patut dipertimbangkan. Teknik regularisasi seperti drop out, L2 regularization, atau *early stopping* dapat membantu mengurangi risiko *overfitting* pada kelas dengan data terbatas. Dengan mengintegrasikan strategi-strategi ini, diharapkan performa model dalam mengenali kelas minoritas pada dataset yang tidak seimbang dapat ditingkatkan secara signifikan.

DAFTAR PUSTAKA

- [1] A. Roihan, P. A. Sunarya, and A. S. Rafika, “Pemanfaatan Machine Learning dalam Berbagai Bidang: Review paper,” *IJCIT Indones. J. Comput. Inf. Technol.*, vol. 5, no. 1, May 2020, doi: 10.31294/ijcit.v5i1.7951.

- [2] E. T. Cahyani, R. Goejantoro, and M. Siringoringo, "Analisis Regresi Logistik Multinomial Bayes untuk Pemodelan Minat Peserta Didik MAN 2 Samarinda Tahun Ajaran 2018/2019," *EKSPONENSIAL*, vol. 13, no. 1, p. 1, Jun. 2022, doi: 10.30872/eksponensial.v13i1.874.
- [3] M. Tulong, C. Mongi, and M. Mananohas, "Regresi Logistik Multinomial Untuk Menentukan Faktor-Faktor Yang Mempengaruhi Pilihan Perguruan Tinggi Pada Siswa SMA dan SMK di Pulau Karakelang Kabupaten Kepulauan Talaud," *d'CARTESIAN*, vol. 7, no. 2, p. 90, Jul. 2018, doi: 10.35799/dc.7.2.2018.21456.
- [4] S. Sonang, A. T. Purba, and S. Sirait, "Prediksi Prestasi Mahasiswa Dengan Menggunakan Algoritma Backpropagation," *J. Tek. Inf. Dan Komput. Teknikom*, vol. 5, no. 1, p. 67, Jun. 2022, doi: 10.37600/tekinkom.v5i1.512.
- [5] E. Setyowati and S. Mariani, "Penerapan Jaringan Syaraf Tiruan dengan Metode Learning Vector Quantization (LVQ) untuk Klasifikasi Penyakit Infeksi Saluran Pernapasan Akut (ISPA)," vol. 4, 2021.
- [6] K. W. Wan *et al.*, "Evaluation of the performance of traditional machine learning algorithms, convolutional neural network and AutoML Vision in ultrasound breast lesions classification: a comparative study," *Quant. Imaging Med. Surg.*, vol. 11, no. 4, pp. 1381–1393, Apr. 2021, doi: 10.21037/qims-20-922.
- [7] T. Si, J. Bagchi, and P. B.C. Miranda, "Artificial Neural Network training using metaheuristics for medical data classification: An experimental study," May 2022.
- [8] Y. Apriyani, "Penerapan Jaringan Syaraf Tiruan Backpropagation Untuk Prediksi Nilai UN Siswa SMPN 2 Cihaurbeuti".
- [9] M. Ikhsan, A. Armansyah, and A. A. Tamba, "Implementasi Jaringan Syaraf Tiruan Backpropagation Pada Klasifikasi Grade Teh Hitam," *J. Sist. Komput. Dan Inform. JSON*, vol. 4, no. 2, p. 387, Dec. 2022, doi: 10.30865/json.v4i2.5312.
- [10] N. Hamdani, A. Setyanto, and S. Sudarmawan, "Perbandingan Algoritma Regresi Logistic Dan Neural Network Pada Prediksi Nilai Hasil Pembinaan Dan Kelulusan Tepat Waktu," *Respati*, vol. 15, no. 1, p. 30, Mar. 2020, doi: 10.35842/jtir.v15i1.328.
- [11] D. Makatulung and A. L. Samal, "Manajemen Waktu Bagi Mahasiswa Hampir Drop Out di IAIN Manado," *J. Islam. Educ. Leadersh.*, vol. 1, no. 2, pp. 136–152, Jan. 2022, doi: 10.30984/jmpi.v1i2.157.
- [12] D. R. N. Fauziyah, "Analisis Data Menggunakan Multiple Logistic Regression Test di Bidang Kesehatan Masyarakat dan Klinis," Apr. 2019.
- [13] M. Fahmuddin, M. K. Aidid, and M. J. Taslim, "Implementasi Analisis Regresi Logistik dengan Metode Machine Learning untuk Mengklasifikasi Berita di Indonesia," Nov. 2023.
- [14] E. Komariah, B. C. Octariadi, and A. C. Siregar, "Penerapan Jaringan Syaraf Tiruan untuk Memprediksi Hasil Produksi Kelapa Sawit Menggunakan Backpropagation," *Jutisi J. Ilm. Tek. Inform. Dan Sist. Inf.*, vol. 12, no. 2, p. 484, Aug. 2023, doi: 10.35889/jutisi.v12i2.1342.
- [15] R. S. Tantika and A. Kudus, "Penggunaan Metode Support Vector Machine Klasifikasi Multiclass pada Data Pasien Penyakit Tiroid," *Bdg. Conf. Ser. Stat.*, vol. 2, no. 2, pp. 159–166, Jul. 2022, doi: 10.29313/bcss.v2i2.3590.
- [16] S. Andjani Musofwan, "Metode Penelitian, Repository.upi.edu." Repository UPI, Aug. 15, 2022.
- [17] G. A. B. Suryanegara, A. Adiwijaya, and M. D. Purbolaksono, "Peningkatan Hasil Klasifikasi pada Algoritma Random Forest untuk Deteksi Pasien Penderita Diabetes Menggunakan Metode Normalisasi," *J. RESTI Rekayasa Sist. Dan Teknol. Inf.*, vol. 5, no. 1, pp. 114–122, Feb. 2021, doi: 10.29207/resti.v5i1.2880.