

KLASIFIKASI KUALITAS AIR YANG LAYAK PAKAI DENGAN *MACHINE LEARNING*

Gabriella Ignatia

Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Tarumanagara,
Jln. Letjen S. Parman No. 1, Jakarta, 11440, Indonesia
E-mail: gabriella.535220104@stu.untar.ac.id

ABSTRAK

Kualitas air berperan penting dalam kesejahteraan dan kesehatan Masyarakat. Untuk mengetahui kualitas air yang baik, diperlukan proses klasifikasi terhadap karakteristik air. Metode yang digunakan, yaitu *K-Nearest Neighbors* (KNN) dan *Artificial Neural Networks* (ANN). Hasil eksperimen kemudian dievaluasi untuk menentukan metode terbaik dalam implementasi klasifikasi terhadap kualitas air yang layak pakai.

Kata kunci klasifikasi, *Machine Learning*, *K-Nearest Neighbors* (KNN), *Artificial Neural Networks* (ANN)

ABSTRACT

Water quality plays a crucial role in the wellbeing and health of the society. In order to assess water with a good water quality classification process is required based on water characteristics. The methods used in this research are K-Nearest Neighbors (KNN) and Artificial Neural Networks (ANN). The experimental results are then evaluated to determine which is the best method for the classification of potable water quality.

Keywords—*Classification, Machine Learning, K-Nearest Neighbors (KNN), Artificial Neural Networks (ANN)*

1. PENDAHULUAN

Air merupakan salah satu sumber daya alam utama bagi kehidupan manusia. Hampir seluruh aktivitas manusia memerlukan air, seperti masak, mencuci, dan minum. Kualitas air jelas berperan penting dalam kehidupan manusia. Air memiliki beberapa karakteristik, meliputi karakteristik fisika dan kimiawi. Indikator air layak untuk digunakan adalah jernih, tidak memiliki bau, bebas dari unsur kimia berbahaya dan bakteri [1]. Kebutuhan air yang layak pakai jelas sangat krusial bagi kehidupan manusia. Kualitas air yang buruk dapat mengakibatkan berbagai penyakit, seperti penyakit kulit. Infeksi pencernaan, dan lainnya [2].

Di era modern ini, Masyarakat sudah tidak asing lagi dengan istilah *Artificial Intelligence*, atau yang sering dikenal dengan istilah AI, yaitu suatu sistem yang diprogram untuk berpikir dan memiliki perilaku layaknya manusia [3]. *Artificial Intelligence* memiliki beberapa cakupan teknik dan pendekatan, seperti *Machine Learning*, *Deep Learning*, dan *Computer Vision* [3]. *Machine Learning* mempelajari pola dari data yang telah ada kemudian menghasilkan suatu model atau algoritma [4-8]. Model atau algoritma hasil *Machine Learning* dapat digunakan untuk membuat model prediksi, klasifikasi, dan lainnya. *Machine Learning* terdiri dari beberapa bagian [4-8], yaitu

Supervised learning, yaitu pendekatan yang melibatkan data yang sudah diberi anotasi atau label, seperti data hasil pengukuran dan data hasil diagnosa. *Unsupervised learning*, yaitu pendekatan yang melibatkan data tanpa label, seperti data sensor. *Semi Supervised learning*, yaitu gabungan antara *supervised* dan *unsupervised* Dimana hanya sebagian data yang diberikan label. *Reinforcement learning*, yaitu proses pembelajaran mesin dengan melakukan percobaan terus menerus (*trial and error*) untuk memperoleh hasil.

Model *supervised learning* dilatih untuk menemukan dan mempelajari pola yang terdapat dalam sumber data [6]. Kemudian, *supervised learning* akan diuji untuk melakukan prediksi atau klasifikasi terhadap data baru berdasarkan pola-pola yang telah ditentukan dari data latih sebelumnya. *Supervised learning* meliputi *Linear Regression*, *K-Nearest Neighbors* (KNN), *Support Vector Classifiers*, *Decision Tree* (DT), *Artificial Neural Network* (ANN) [6,7]. Klasifikasi adalah proses pengelompokan objek dengan karakteristik yang sama berdasarkan batasan yang ditentukan [9]. Algoritma K-Nearest Neighbors (K-NN) melakukan klasifikasi dari *data training* atau dataset yang sudah ada berdasarkan kelompok k objek yang memiliki karakteristik mirip dengan data baru atau *data testing* [10]. Algoritma (ANN) meniru cara kerja perpindahan saraf pada jaringan sistem saraf manusia dimana data *input* dapat berupa data mentah atau hasil proses neuron sebelumnya dan data *output* dapat berupa hasil akhir atau data *input* untuk neuron selanjutnya [11].

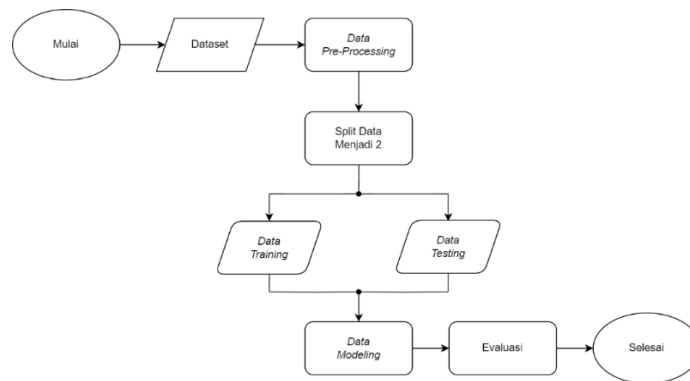
Dengan memanfaatkan *Machine Learning*, diharapkan proses klasifikasi dapat dilakukan untuk memperoleh kualitas air yang layak pakai. Kualitas air yang baik mempengaruhi peningkatan kesejahteraan dan kesehatan masyarakat.

2. METODE PENELITIAN

Dataset untuk klasifikasi ini diperoleh dari situs Kaggle (Gambar 1). Dataset ini memiliki 8000 sampel dan 21 kolom, 20 kolom merupakan daftar kandungan dalam air dan 1 kolom mengenai kelayakan dari air tersebut. Klasifikasi ini dilakukan menggunakan algoritma *K-Nearest Neighbors* (K-NN) dan *Artificial Neural Network* (ANN) dengan prose seperti *flowchart* pada Gambar 2.

	# aluminum	# ammonia	# arsenic	# barium	# cadmium	# chloramine	# chromium	# copper	# fluoride	# bacteria	# viruses	# lead	# nitrate	# nitrite	# mercury	# pesticides	# radium	# selenium	# silver	# uranium	# sulfate	
	1.05	0.00	0.04	2.05	0.007	0.35	0.03	0.17	0.05	0.2	0	0.054	16.00	1.13	0.007	37.75	0.70	0.00	0.34	0.02	1	
	2.32	21.16	0.01	3.31	0.002	5.39	0.00	0.00	0.9	0.05	0.05	0.1	2.01	1.93	0.003	32.26	3.21	0.00	0.27	0.05	1	
	1.01	14.02	0.04	0.50	0.000	4.24	0.53	0.02	0.99	0.05	0.003	0.070	14.16	1.11	0.006	50.39	7.07	0.07	0.44	0.01	0	
	1.36	11.33	0.04	2.96	0.001	7.23	0.03	1.00	1.00	0.71	0.71	0.016	1.40	1.29	0.004	0.12	1.72	0.02	0.45	0.05	1	
	0.92	34.33	0.03	0.2	0.006	2.07	0.09	0.57	0.01	0.13	0.001	0.117	0.74	1.11	0.003	16.9	2.41	0.02	0.06	0.02	1	
	0.94	14.47	0.03	2.00	0.003	0.0	0.43	1.30	0.11	0.07	0.135	0.75	1.00	0.006	27.17	5.42	0.00	0.19	0.02	1		
	2.36	5.6	0.01	1.35	0.004	1.30	0.02	1.00	0.33	0.13	0.007	0.021	10.6	1.70	0.007	45.34	2.04	0.1	0.24	0.00	0	
	3.93	19.07	0.04	0.66	0.001	0.22	0.1	1.00	0.06	0.16	0.005	0.107	13.05	1.01	0.001	53.35	7.24	0.00	0.00	0.07	0	
	0.6	34.50	0.01	0.71	0.005	3.14	0.77	1.45	0.90	0.35	0.002	0.107	14.60	1.04	0.004	23.45	4.99	0.00	0.25	0.00	1	
	0.22	16.70	0.02	1.37	0.007	0.4	0.49	0.02	1.24	0.03	0.03	0.100	4.70	1.46	0.01	30.42	0.00	0.03	0.31	0.01	1	
	2.27	3.6	0.001	2.69	0.005	5.75	0.15	0.6	1.29	0.04	0.000	0.145	0.47	1.25	0.006	55.4	0	0.05	0.33	0.06	0	
	1.35	21.96	0.04	0.04	0.002	0.1	0.76	0.17	0.50	0.52	0.011	10.4	1.49	0.000	21.52	1.3	0.00	0.40	0.00	1		
	1.00	19.20	0.02	2.70	0.000	0.05	0.42	1	0.00	0.01	0.91	0.103	4.37	1.95	0.006	22.12	1.07	0.03	0.06	0.05	1	
	4.93	23.90	0.04	3.05	0.000	0.7	0.51	1.35	1.07	0.7	0.7	0.101	1.16	1.11	0.000	26.0	5.50	0.00	0.30	0.03	1	
	2.09	10.02	0.05	3.77	0.000	5.99	0.54	0.79	0.54	0.2	0.009	0.126	17.50	1.02	0.009	17.54	4.23	0.1	0.05	0.02	1	
	0.01	2.41	0.03	0.50	0.002	1.94	0.77	1.50	0.02	0.001	0.017	1.00	1.00	0.007	11.16	0.90	0.01	0.47	0.03	1		
	2.47	15.04	0.02	0.06	0.001	5.29	0.47	1.00	1.45	0.00	0.00	0.00	1.91	1.2	0.000	0.10	0.00	0.06	0.12	0.00	1	
	2.11	17.03	0.02	0.00	0.000	7.70	0.00	1.15	0.34	0.05	0.05	0.005	17.06	1.53	0.003	19.4	1.14	0.1	0.4	0.01	1	
	4.00	26.94	0.02	0.30	0.001	1.21	0.60	0.71	0.00	0.75	0.75	0.071	0.31	1.22	0.002	50.7	1	0	0.41	0.05	0	
	4.12	17.00	0.02	3.43	0.000	0.01	0.41	1.02	0.22	0.00	0.99	0.99	0.100	0.00	1.70	0.005	24.29	0.00	0.1	0.1	0.07	1
	0.00	10.99	0.001	0.04	0.000	4.57	0.2	1.10	1	0.92	0.92	0.006	9.40	1.41	0.007	21.70	3.05	0.03	0.13	0.00	1	
	1.15	0.12	0.02	0.07	0.007	3.47	0.05	1.51	1.40	0.00	0.00	0.001	0.96	1.5	0.004	14.6	1.74	0.03	0.01	0.06	1	
	0.27	10.67	0.02	0.55	0.001	3.74	0.12	1.77	0.45	0.0	0.0	0.114	12.69	1.10	0.000	34.64	0.9	0.02	0.10	0.06	1	
	4.32	20.64	0.03	2.6	0.000	7.24	0.01	1.23	1.44	0.50	0.50	0.012	9.42	1.74	0.004	50.23	3.22	0.07	0.10	0.00	0	
	2.36	27.05	0.01	0.60	0.003	4.07	0.13	1.34	0.29	0.06	0.96	0.107	15.05	1.92	0.002	50.32	7.00	0.06	0.5	0.06	0	
	3.31	22.07	0.03	0.40	0.001	7.22	0.73	1.05	1	0.25	0.007	0.109	1.92	1.07	0.001	39.4	0.40	0.04	0.47	0.05	1	
	1.02	0.01	0.01	0.05	0.000	2.55	0.25	1.00	1.35	0.10	0.002	0.031	16.99	1.7	0.007	44.76	1.15	0.00	0.20	0.00	1	
	3.42	2.4	0.001	2.0	0.003	2.07	0.73	0.27	0.53	0.44	0.002	0.11	13.73	1.00	0.000	55.27	3.29	0.04	0.17	0	0	
	4.41	15.14	0.03	1.70	0.007	0.03	0.02	1.57	0.26	0.00	0.09	0.102	1.40	1.01	0.000	24.91	2.30	0.01	0.2	0.05	1	
	4.57	25.04	0.01	3.04	0.002	2.70	0.72	0.40	1.41	0.00	0.000	0.049	9.04	1.43	0.000	50.57	4.55	0.03	0.19	0	0	
	1.09	22.0	0.001	2.01	0.002	5.07	0.70	1.53	0.07	0.13	0.000	0.070	0.04	1.19	0.007	24.01	7.77	0.06	0.11	0.05	1	
	1.07	0.1	0.05	3.74	0.000	2.06	0.52	1.03	0.51	0.77	0.77	0.012	1.11	1.55	0.000	20.32	2.21	0.02	0.40	0.04	1	
	2.71	26.19	0.04	3.02	0.000	4	0.17	1.50	0.91	0.30	0.005	0.013	11	1	0.000	14.16	0.10	0.05	0.05	0.00	1	
	1.03	15.75	0.03	2.54	0.000	4.25	0.74	1.35	1.24	0.12	0.000	0.103	17.20	1.27	0.001	50.15	1.71	0.00	0.44	0.05	0	
	0.01	20.20	0.001	2.93	0.007	7.75	0.00	1.00	0	0.0	0.0	0.001	17.33	1.00	0.001	43.00	3.70	0.00	0.3	0	0	
	4.00	4.07	0.05	1.07	0.005	4.55	0.32	0.9	0.55	0.3	0.002	0.135	9.23	1.43	0.004	50.32	0.15	0.04	0.10	0.04	1	

Gambar 1 Tabel Dataset



Gambar 2 Flowchart Penelitian

2.1 Data Pre-processing

Data pre-processing perlu untuk dilakukan sebelum data diolah untuk menghindari kesalahan pada tahap pengolahan [12]. Kesalahan yang terjadi dapat berupa *missing value*, duplikasi data, imbalansi, dan lainnya. Ketidakseimbangan data dapat mengakibatkan model lebih bias ke satu kelas daripada kelas lain yang minoritas [13]. Tentunya hal ini akan mengganggu proses pengolahan data sehingga perlu dilakukan penyeimbangan data.

2.2 Data Modeling

Setelah proses penyeimbangan data, dataset sudah siap untuk diolah. Untuk pemodelan data pada klasifikasi ini, digunakan dua algoritma, yaitu *K-Nearest Neighbors* (K-NN) dan *Artificial Neural Network* (ANN).

2.2.1 K-Nearest Neighbors (K-NN)

K-Nearest Neighbors melakukan klasifikasi berdasarkan data latih dengan jarak terdekat terhadap suatu objek [14]. Sebelum melakukan klasifikasi, K-NN menentukan nilai K tetangga. Lalu untuk memperoleh jarak terdekat, digunakan persamaan sebagai berikut (Persamaan 1) [15-16].

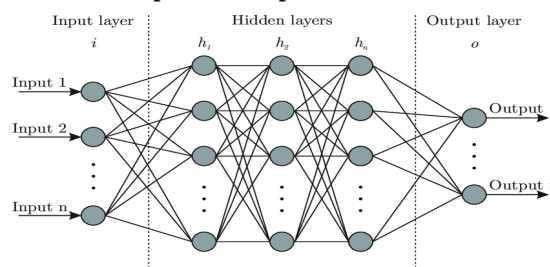
$$d(x, y) = \sqrt{\sum_{i=1}^m (x_i - y_i)^2} \quad (1)$$

Keterangan:

$d(x, y)$: Jarak antara dua titik
 x : Data latih
 y : Data uji
 m : Dimensi data
 i : Variabel data

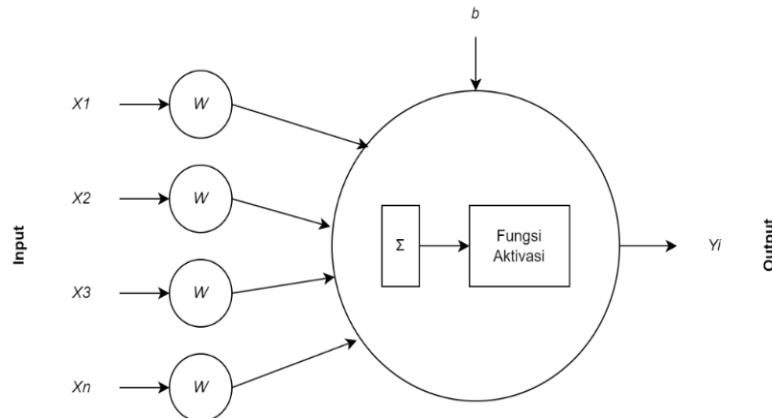
2.2.2 Artificial Neural Network (ANN)

Artificial Neural Network (ANN) memproses data menggunakan cara perpindahan rangsangan saraf pada jaringan sistem saraf manusia [11, 17]. ANN memproses data kemudian diteruskan ke neuron selanjutnya. Arsitektur ANN dapat dilihat pada Gambar 3.



Gambar 3 Arsitektur ANN

Secara umum, arsitektur ANN memiliki tiga lapisan (*layer*), yaitu sebagai berikut [18]. *Input Layer*, lapisan awal dengan bobot bergantung pada banyaknya dimensi input. *Hidden Layer*, lapisan tempat dilakukannya proses perhitungan. Makin kompleks jaringan, makin banyak jumlah lapisan yang tersembunyi. *Output Layer*, lapisan terakhir yang meneruskan hasil perhitungan lapisan tersembunyi sebagai *output*. Banyaknya dimensi lapisan ini bergantung pada dimensi *output*. Skema konfigurasi *node* pada ANN dapat dilihat pada Gambar 4.



Gambar 4 Skema Konfigurasi ANN

Berdasarkan Gambar 5, diperoleh perhitungan seperti yang tertera pada persamaan 2.

$$y = a \left(\sum_{i=1}^n w_i x_i + b \right) \quad (2)$$

Keterangan:

y : Nilai *output*
 a : Nilai fungsi aktivasi
 w : Nilai bobot
 x : Nilai *input*
 b : Nilai bias
 i : Variabel data

2.3 Evaluasi Model

Evaluasi model perlu dilakukan untuk menentukan tingkat akurasi hasil klasifikasi yang telah dilakukan [19]. Evaluasi model dilakukan dengan menggunakan *confusion matrix*.

2.3.1 Confusion Matrix

Confusion matrix menunjukkan tingkat akurasi, presisi, dan *recall* dari hasil klasifikasi yang telah dilakukan [20]. Perhitungan *confusion matrix* dapat dilihat pada Tabel 1.

Tabel 1 Perhitungan *Confusion Matrix*

Kelas Sebenarnya	Kelas Prediksi		
		P	N
	P	TP	FN
	N	FP	TN

Berdasarkan tabel di atas, terdapat empat hasil perhitungan *confusion matrix*, yaitu.

- *True Positive* (TP) – model mengidentifikasi kelas positif dengan benar.
- *True Negative* (TN) – model mengidentifikasi kelas negative dengan benar.
- *False Positive* (FP) – model salah mengidentifikasi kelas positif, seharusnya termasuk ke dalam kelas negatif.
- *False Negative* (FN) – model salah mengidentifikasi kelas negatif, seharusnya termasuk ke dalam kelas positif.

Setelah diperoleh hasil *confusion matrix*, selanjutnya dilakukan perhitungan untuk mendapatkan nilai *accuracy*, *precision*, *f1-score* dan *recall*, seperti yang dapat dilihat pada persamaan (3), (4), (5), dan (6).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (3)$$

$$Precision = \frac{TP}{TP + FP} \quad (4)$$

$$f1 - score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (5)$$

$$Recall = \frac{TP}{TP + FN} \quad (6)$$

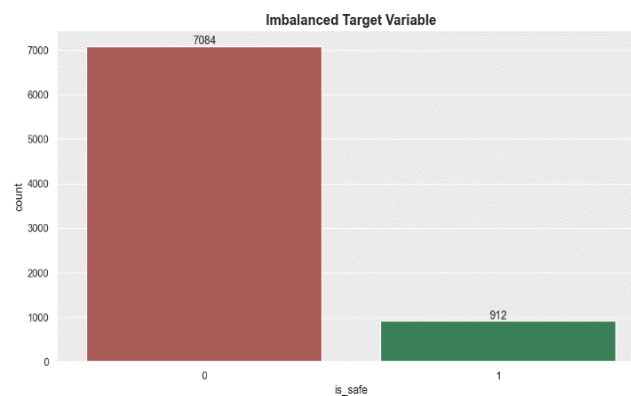
3. HASIL DAN PEMBAHASAN

Proses klasifikasi ini dijalankan dengan menggunakan bahasa pemrograman Python. Dilakukan *data pre-processing* terlebih dahulu untuk memeriksa kesiapan dan kelengkapan dataset. Gambar 5 menunjukkan informasi dari dataset.

Diketahui bahwa tidak ada *missing value* dan duplikasi pada dataset. Namun, terdapat ketidakseimbangan data (Gambar 6). Ini tentunya akan mengganggu proses pengolahan data, sehingga dataset harus diseimbangkan terlebih dahulu.

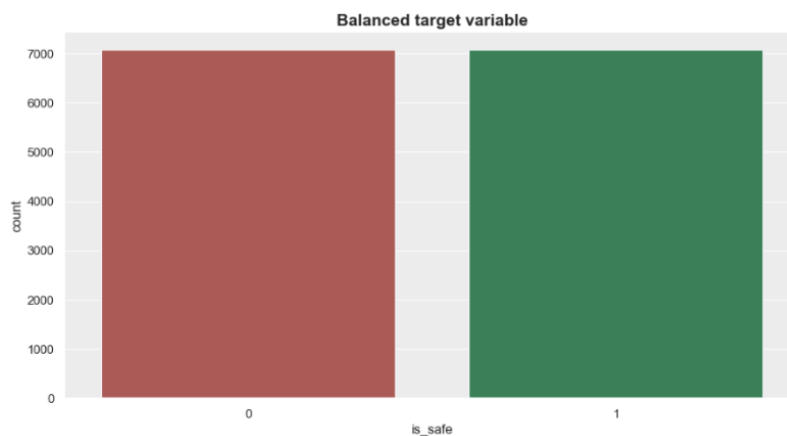
```
RangeIndex: 7999 entries, 0 to 7998
Data columns (total 21 columns):
#   Column          Non-Null Count  Dtype
---  -
0   aluminium       7999 non-null   float64
1   ammonia         7999 non-null   object
2   arsenic         7999 non-null   float64
3   barium          7999 non-null   float64
4   cadmium         7999 non-null   float64
5   chloramine      7999 non-null   float64
6   chromium        7999 non-null   float64
7   copper          7999 non-null   float64
8   flouride        7999 non-null   float64
9   bacteria        7999 non-null   float64
10  viruses         7999 non-null   float64
11  lead            7999 non-null   float64
12  nitrates        7999 non-null   float64
13  nitrites        7999 non-null   float64
14  mercury         7999 non-null   float64
15  perchlorate     7999 non-null   float64
16  radium          7999 non-null   float64
17  selenium        7999 non-null   float64
18  silver          7999 non-null   float64
19  uranium         7999 non-null   float64
20  is_safe         7999 non-null   object
dtypes: float64(19), object(2)
memory usage: 1.3+ MB
```

Gambar 5 Informasi Dataset



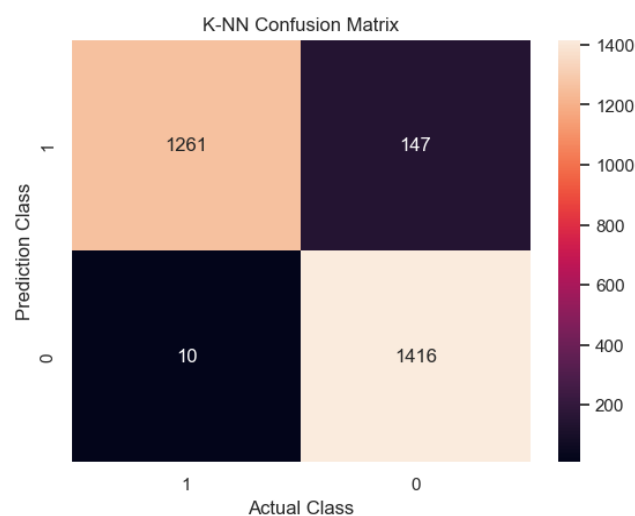
Gambar 6 *Imbalanced Target Variable*

Penyeimbangan data dilakukan dengan implementasi kelas *RandomOverSampler* pada jumlah baris yang sama untuk nilai 1 dengan tujuan mendapatkan jumlah baris yang sama untuk kedua nilai target. Hasil penyeimbangan dapat dilihat pada Gambar 7.

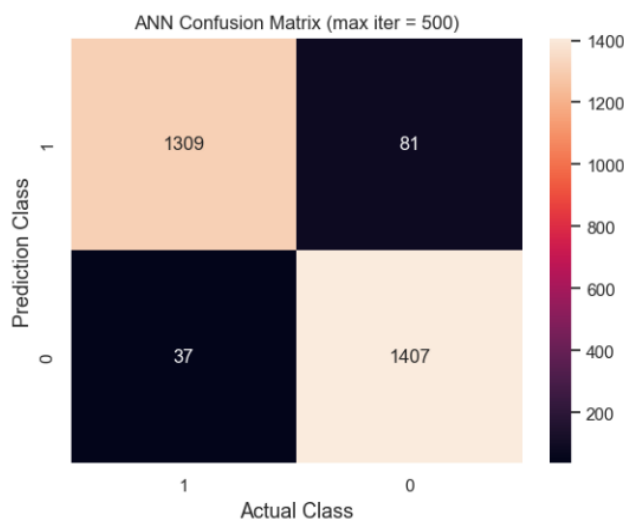


Gambar 6 *Balanced Target Variable*

Setelah proses penyeimbangan, dataset sudah siap untuk diproses. Dataset dibagi menjadi data latih dan data uji dengan perbandingan 80:20. Kemudian, *data modeling* dilakukan dengan algoritma K-NN dan ANN. Hasil *confusion matrix* terdapat pada Gambar 7 dan 8.



Gambar 7 *K-NN Confusion Matrix*



Gambar 8 ANN Confusion Matrix

Setelah diperoleh hasil dari *confusion matrix*, selanjutnya akan dilakukan perhitungan untuk nilai *accuracy*, *precision*, *f1-score*, dan *recall*. Hasil perhitungan ditampilkan pada Tabel 2.

Tabel 2 Hasil Klasifikasi

	Algoritma	
	K-NN	ANN
<i>Accuracy</i>	94,53%	95,84%
<i>Precision</i>	90,55%	94,56%
<i>Recall</i>	94%	96%
<i>F1-score</i>	95%	96%

4. KESIMPULAN

Berdasarkan klasifikasi yang telah dilakukan, tingkat akurasi menggunakan algoritma K-NN diperoleh sebesar 94,53% dan tingkat akurasi menggunakan algoritma ANN sebesar 95,84%. Dengan ini, dapat disimpulkan bahwa hasil klasifikasi dataset dengan algoritma ANN lebih akurat.

UCAPAN TERIMA KASIH

Dengan diselesaikannya penyusunan jurnal Klasifikasi Kualitas Air yang Layak Pakai dengan *Machine Learning*, ucapan syukur dipanjatkan kepada Tuhan Yang Maha Esa juga ucapan terima kasih kepada Dosen Mata Kuliah *Machine Learning*, Bu Teny Handayani dan seluruh teman serta pihak lainnya yang ikut memberikan dukungan selama proses penyusunan jurnal ini.

DAFTAR PUSTAKA

- [1] Wicaksono, B., Iduwin, T., Mayasari, D., Putri, P. S., Yuhanah, T., 2019, Edukasi Alat Penjernih Air Sederhana Sebagai Upaya Pemenuhan Kebutuhan Air Bersih, *Jurnal Pengabdian Pada Masyarakat Menerangi Negeri (TERANG)*, No. 1, Vol. 2, hal. 1-82.
- [2] Rosdiana, D., Hastiaty, I. A., Hartomy, E., Kango, I., Simbolon, P. T., Pradapaningrum, P. G., Indriasih, M., Paramasatya, A., 2023, Kontaminasi Kimia dan Biologi pada Air dan Udara dengan ARKM: Analisis Risiko Kesehatan Masyarakat, *Public Health Risk Assessment Journal*, No. 1, Vol. 1, hal.1-20.
- [3] Nawi, A., 2019, Penerokaan Awal terhadap Isu dan Impak Penggunaan Teknologi Kecerdasan Buatan terhadap Kehidupan Manusia. *Asian Journal of Civilization Studies (AJOCS)*, No. 4, Vol. 1, hal. 24-33.

- [4] Pakpahan, R., 2021, Analisa Pengaruh Implementasi Artificial Intelligence dalam Kehidupan Manusia, *Journal of Information System, Informatics and Computing (JISICOM)*, No. 2, Vol. 5, hal. 506-513.
- [5] Roihan, A., Sunarya, P. A., Rafika, A. S., 2020, Pemanfaatan Machine Learning dalam Berbagai Bidang: Review Paper, *Indonesian Journal on Computer and Information Technology (IJCIT)*, No. 1, Vol. 5, hal. 75-82.
- [6] Mestika, J. C., Selan, M. O., Qadafi, M. I., 2022, Menjelajahi Teknik-teknik Supervised Learning untuk Pemodelan Prediktif Menggunakan Python, *Buletin Ilmiah Ilmu Komputer dan Multimedia (BIIKMA)*, No. 99, Vol. 99, hal. 216-219.
- [7] Bi, G., Goodman, K. E., Kaminsky, J., Lessler, J., 2019, What is Machine Learning? A Primer for the Epidemiologist, *American Journal of Epidemiology*, No. 12, Vol. 188, hal. 2222-2239.
- [8] Khan, A. A., Laghari, A. A., Awan, S. A., 2021, Machine Learning in Computer Vision: A Review, *Endorsed Transactions on Scalable Information Systems (EAI)*, No. 32, Vol. 8.
- [9] Sidik, A. D. W. M., Edwinanto, Kusumah, I. H., Artiyasa, M., Suryana, A., Junfithrana, A., P., 2020, Gambaran Umum Metode Klasifikasi Data Mining, *Jurnal Teknik Elektro (FIDELITY)*, 202f. 0, No. 2, Vol; 2, hal. 34-38.
- [10] Admojo, F. T., Ahsanawati, 2020, Klasifikasi Aroma Alkohol Menggunakan Metode KNN, *Indonesian Journal of Data and Science (IJODAS)*, No. 2, Vol. 1, hal. 34-38.
- [11] Pradana, D. G., Alghifari, M. L., Juna, M. F., Palaguna, S. D., 2022, Klasifikasi Penyakit Jantung Menggunakan Metode Artificial Neural Network, *Indonesian Journal of Data and Science (IJODAS)*, No. 2, Vol. 3, hal. 55-60.
- [12] Maharana, K., Mondal, S., Nemade, B., 2022, A Review: Data Pre-processing and Data Augmentation Techniques, *Global Transitions Proceedings*, No. 1, Vol. 3, hal. 91-99.
- [13] Ali, H., Salleh, M. N. M., Saendudin, R., Husain, K., Mushtaq, M. F., 2019, Imbalance Class Problem in Data Mining: A Review, *Indonesian Journal of Electrical Engineering and Computer Science*, No. 3, Vol. 14, hal. 1560-1571.
- [14] Manullang, R. A., Sianturi, F. A., 2021, Penerapan Algoritma K-Nearest Neighbour untuk Memprediksi Kelulusan Mahasiswa, *Jurnal Ilmu Komputer dan Sistem Informasi (JIKOMSI)*, No. 2, Vol. 4, hal. 42-50.
- [15] Prakoso, R. D. Y., Wiriaatmadja, B. S., Wibowo F. W., 2020, Sistem Klasifikasi pada Penyakit Parkinson Menggunakan Metode K-Nearest Neighbor, *Seminar Nasional Teknologi Komputer & Sains (SAINTEKS)*, HAL. 63-68.
- [16] Qiudandra, E., Akram, R., Novianda, 2022, Sistem Pakar Diagnosa Penyakit Osteoarthritis dengan Menggunakan Metode K-Nearest Neighbor, *Jurnal Ilmiah Teknik Informatika (METHOTHIKA)*, No. 2, Vol. 2, hal. 37-48.
- [17] Simajuntak, H. T. A., Sigiyo, M., Simanungkalit, H. M., 2021, Deteksi Fraud pada Klaim Layanan Rumah Sakit Menggunakan Model Neural Netwok, *Journal of Applied Technology and Informatics Indonesia*, No. 1, Vol. 1.
- [18] Jaddoa, A. S., Saba, S. J., Al-Kareem, E. A., 2023, Liver Prediction Model Based on Oversampling Dataset with RFE Feature Selection Using ANN and AdaBoost Algorithms, *Buana Information Technology and Computer Sciences (BIT and CS)*, No. 2, Vol. 4, hal. 85-93.
- [19] Tarigan, I., D., S., Habibi, R., Fatonah, Rd. N. S., 2023, Evaluasi Algoritma Klasifikasi Machine Learning Kategori Nilai Akhir Tugasan Kinerja Pegawai, *Jurnal Sistem Cerdas*, No. 3, Vol. 6, hal. 251-261.
- [20] Normawati, D., dan Prayogi, S. A., 2021, Implementasi Naïve Bayes Classifier dan Confusion Matrix pada Analisis Sentimen Berbasis Teks pada Twitter, *Jurnal Sains Komputer dan Informatika (J-SAKTI)*, No. 2, Vol. 5, hal. 697-711.