Jurnal Komputer dan Informatika
Vol 19 No 2, Oktober 2024: hlm 139-147

ISSN 1410-7228 (Versi Cetak) ISSN
2716-4233 (Versi Elektronik)

KLASIFIKASI INDEKS STANDAR PENCEMARAN UDARA DI JAKARTA DENGAN METODE *K-NEAREST NEIGHBORS* DAN *ARTIFICIAL NEURAL NETWORK*

Nicholas Eugene Supardi
Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Tarumanagara,
Jln. Letjen S. Parman No. 1, Jakarta, 11440, Indonesia
E-mail:Nicholas.535220102@stu.untar.ac.id

ABSTRAK

Penelitian ini melihat bagaimana dua algoritma machine learning, *K-Nearest Neighbo*r (KNN) dan *Artificial Neural Network* (ANN), berfungsi untuk memprediksi Indeks Standar Pencemaran Udara (ISPU) di Jakarta. Data yang digunakan berasal dari tahun 2014 hingga 2019, dan mencakup sepuluh fitur, termasuk kelas target dan parameter kualitas udara yang didasarkan pada indeks pencemaran udara. Setelah tahap pembersihan dan pembagian data, kedua algoritma dilatih dan diuji berdasarkan metrik akurasi, presisi, recall, dan skor F-1. Hasilnya menunjukkan bahwa KNN memiliki akurasi 98%, sedangkan ANN memiliki akurasi 97%. Dengan demikian, KNN lebih cocok untuk memprediksi dataset ISPU Jakarta. Penelitian ini memberikan wawasan tentang bagaimana algoritma pembelajaran mesin dapat digunakan untuk mengatasi masalah kualitas udara di perkotaan.

Kata kunci—*K-Nearest Neighbor, Artificial Neural Network,* Indeks Standar Pencemaran Udara, Jakarta, *Machine Learning.*

ABSTRACT

This research looks at how two machine learning algorithms, K-Nearest Neighbor (KNN) and Artificial Neural Network (ANN), function to predict the Air Pollution Standard Index (ISPU) in Jakarta. The data used is from 2014 to 2019 and includes ten features, including target classes and air quality parameters that are based on ISPU. After the data cleaning and sharing stage, both algorithms were trained and tested based on accuracy, precision, recall, and F-1 score metrics. The results show that KNN has 98% accuracy, while ANN has 97% accuracy. Thus, KNN is more suitable for predicting the Jakarta ISPU dataset. This research provides insight into how on those algorithms can be used to address urban air quality issues.

Keywords— *K-Nearest Neighbor, Artificial Neural Network, Air Pollution Standard Index, Jakarta, Machine Learning.*

1. PENDAHULUAN

Masalah kualitas udara, selalu berkaitan dengan perkembangan kota – kota besar di dunia ini, termasuk di Indonesia. Setiap harinya kualitas udara di kota – kota besar semakin menurun karena tingginya polusi udara, hal ini disebabkan oleh meningkatnya penggunaan kendaraan pribadi, industri, pembakaran, dan penumpukan sampah. Oleh karena itu kualitas udara mengalami perubahan yang sungguh drastis [1]. Salah satu kota yang ada di Indonesia adalah Jakarta, berdasarkan standar AQG (*Air Quality Guide*) yang dibuat oleh WHO (*World Health Organization)* pencemaran udara yang ada di Jakarta sudah melewati batas standar tersebut [2]. Pencemaran udara di Jakarta lebih banyak disebabkan oleh kendaraan bermotor, banyaknya orang yang memiliki kendaraan bermotor pada tahun 2018 – 2020, terdapat peningkatan yang cukup signifikan yaitu sebesar 19 % atau sebanyak 8381000 kendaraan dari tahun sebelumnya [3]. Berdasarkan data diatas, Jakarta memiliki polusi yang dikategorikan sedang. Dimana rentang indeks berada pada nilai 51-100 (µg/m3). Rentang indeks tersebut diukur dari berbagai pengukuran gas yakni PM10, SO2, CO, O3, dan NO2 [2]. Klasifikasi adalah proses untuk menemukan model atau fungsi untuk membedakan dan menggambarkan kelas data atau konsep [4]. Hal ini digunakan untuk memprediksi kelas yang tidak

diketahui dari objek pengamatan. Dalam era ini yaitu era *big data*, Dimana jumlah data yang besar terus bertambah, diperlukan metode analisis yang mampu menangani *big data* secara cepat dan akurat. Salah satu metode yang sedang dikembangkan adalah *machine learning*. *Machine Learning* adalah teknologi kecerdasan buatan Dimana sebuah mesin dapat belajar tanpa arahan dari pengguna [4]. Penelitian ini dibuat untuk membandingkan dua algoritma klasifikasi supervised learning yaitu *K-Nearest Neighbor* (KNN) dan *Artificial Neural Network* (ANN) [6]. *K-Nearest Neighbor* adalah algoritma supervised learning Dimana kelas yang paling banyak muncul (mayoritas) yang akan menjadi hasil klasifikasi [7]. *Artificial Neural Network* adalah algoritma klasifikasi yang memiliki 3 struktur / lapisan yakni, input layer, hidden layer, dan output layer. ada sebuah penghubung yang disebut dengan neuron [8].

Isi dari penelitian ini adalah perbandingan kinerja kedua algoritma klasifikasi, yakni *K-Nearest Neighbor* dan *Artificial Neural Network.* Perbandingan akan dinilai berdasarkan hasil dari metrik evaluasi kedua algoritma. Eksperimen akan dimulai dengan mlakukan tahap *data cleansing* untuk membersihkan data, setelah itu melakukan *data splitting*, dan melakukan pelatihan model dengan kedua algoritma yang menghasilkan metrik evaluasi. Penelitian ini di buat untuk memberikan wawasan tentang algoritma machine learning yaitu KNN dan ANN dalam *dataset* Indeks Standar Pencemaran Udara di wilayah Jakarta.

2. METODE PENELITIAN

2.1 Klasifikasi

Salah satu metode untuk memetakan data ke dalam kelompok yang sudah ditentukan adalah klasifikasi. Klasifikasi adalah metode pembelajaran yang diawasi yang membutuhkan data pelatihan untuk menghasilkan aturan untuk mengklasifikasikan data uji ke dalam kelompok yang sudah ditentukan.

2.2 Data *Collection*

Dataset yang digunakan dalam penelitian ini adalah Data Indeks Standar Pencemaran Udara di Jakarta tahun 2014 – 2019 yang dikumpulkan dari Dinas Lingkungan Hidup DKI Jakarta dari website https://data.jakarta.go.id/. Ada 10 atribut yaitu: tanggal, stasiun, CO, PM, NO, O, SO, maks, kritis, dan kategori [1].

2.3 *Splitting* Data

Data yang digunakan dalam penelitian dapat dibagi menjadi dua atau lebih bagian melalui proses yang dikenal sebagai splitting data. Data set biasanya dibagi menjadi data latih (*training*) dan data uji (*testing*). Data latih digunakan untuk melatih algoritma, sementara data uji digunakan untuk mengevaluasi kinerja algoritma [10].

2.4 Data *Pre-Processing*

Setelah data sudah di dapatkan, kita akan melakukan data *pre-processing*. Pada tahap ini akan dilakukan pembersihan data. dalam proses ini kita juga akan melakukan *splitting data*. tahap ini juga mempersiapkan data untuk melakukan prediksi [11]. Selanjutnya *data cleansing* akan dilakukan. Hal ini dilakukan agar pada saat penelitian tidak ada data yang null atau kosong. Selain itu, pemilihan fitur akan dilakukan untuk menerapkan algoritma. Hal ini dilakukan agar dapat mengurangi kompleksitas attribute yang akan dipelajari algoritma [12].

2.5 *K-Nearest Neighbor (KNN)*

Algoritma KNN adalah algoritma *supervised learning* yang dapat digunakan untuk tugas regresi dan klasifikasi. Algoritma bekerja dengan menemukan titik data K yang paling dekat dengan sample yang diberikan, dan kemudian manganoan label kelas atau nilai dari titik data ini untuk membuat prediksi tentang label kelas atau nilai sampel. Dalam tugas klasifikasi, KNN dapat

ISSN 1410-7228 (Versi Cetak)
ISSN 2716-4233 (Versi Elektronik)

melakukan prosedur berbasis matematika untuk mengevaluasi nilai-nilai kriteria tersebut ke dalam klasifikasi. Setelah itu, sampel dimasukkan ke label kelas yang paling umum di antara algoritma KNN [13].

Untuk mengklasifikasikan sekumpulan data, algoritma *K-Nearest Neighbor* (KNN) ini menggunakan pembelajaran data yang sudah terklasifikasi sebelumnya. termasuk dalam pembelajaran yang diawasi, di mana hasil query instance baru dikategorikan berdasarkan mayoritas jarak dari kategori KNN yang ada saat ini. Metode ini menggunakan jarak terkecil antara sampel uji dan sampel latih untuk menentukan KNN sampel uji. Setelah mengumpulkan KNN, sebagian besar KNN digunakan untuk membuat prediksi dari sampel uji. Untuk mengetahui seberapa dekat atau jauh seseorang dengan tetangganya, metode eucledian distance biasanya digunakan. Untuk menghitung metode K-Nearest Neighbor, langkah-langkah berikut harus dilakukan [14]:

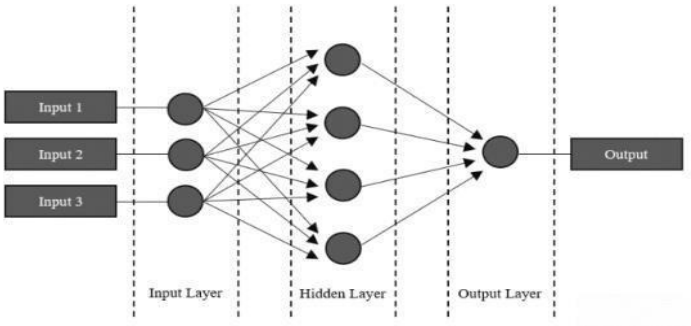
1. Menentukan parameter K
2. Menghitung jarak antara data pelatihan dan data pengujian
   Perhitungan jarak yang paling umum adalah menggunakan perhitungan *Euclidean distance*. Rumus adalah seperti berikut

$$euc = \sqrt{(\sum_{i=1}^{n} (pi - qi)^2)} \quad (1)$$

   Dimana:
   Pi = sample data / *data training*
   Qi = data uji / *data testing*
   i = variabel data
   n = dimensi data

3. Mengurutkan jarak yang terbentuk
4. Menentukan jarak terdekat sampai urutan K.
5. Tempatkan kelas yang sesuai
6. Cari berapa banyak kelas yang ada di tetangga yang terdekat

2.6 *Artificial Neural Network (ANN)*



Gambar 1 Struktur ANN dengan 1 hidden layer

Algoritma *Artificial Neural Network* atau juga dikenal sebagai ANN adalah sebuah algoritma yang dibangun untuk meniru struktur otak manusia dengan tujuan meniru cara otak menanggapi pembelajaran. ANN terdiri dari kumpulan neuron yang saling terhubung, seperti otak manusia. ANN dapat memperkirakan operasi yang sangat kompleks dan non-linear melalui arsitektur yang kompleks. Untuk berhasil dalam berbagai aplikasi seperti estimasi, klasifikasi, dan pengenalan, ANN dapat dilatih secara bertahap dengan memproses data pelatihan dan mengubah parameter model sesuai dengan berbagai metode pelatihan sampai *output* model tercapai [15].

Artificial Neural Network (ANN) bergantung pada model dengan aturan aktivasi, perkalian, dan penjumlahan. Nilai input artificial neuron terdiri dari bobot individual setiap input dikalikan, kemudian penjumlahan bobot dan bias. Setelah fungsi aktivasi selesai, penjumlahan bobot dan bias diteruskan ke *output neural network*. Prinsip kerja dan aturan ANN tampaknya sangat sederhana. Namun, jika ANN terhubung satu sama lain, modelnya menjadi lebih canggih dan dapat dihitung (Gambar 1). Kompleksitas dapat berasal dari sejumlah aturan dasar sederhana [16].

Pada *hidden layer*. terdapat sebuah aktifitas *neuron* yang ditentukan oleh sejumlah input (xk), Dimana k = 1,2,..,i. setelah itu dia melewati *input layer*. proses melewati dari input layer menghasilkan output yang akan di teruskan ke *hidden layer*. output ini dinamakan *synaptic weight* (wkl), Dimana l = 1,2,.., i. dan yang terakhir ada hasil output dari hidden layer (hl) [17].

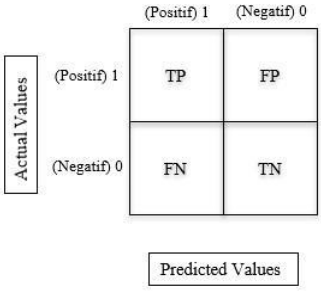
$$hl = f\left(\sum_{k=1}^{i} xk.wk + w0\right) \tag{2}$$

Dengan menggunakan rumus yang sama seperti yang digunakan pada persamaan (2), kita dapat menemukan *neuron* untuk *output* lapisan (ymp), yang dihasilkan dari *output* lapisan tersembunyi dan tekanan *synaptic* antara lapisan tersembunyi dan *output* lapisan (vlm) [15].

$$ymp = f\left(\sum_{l=1}^{i} h1.v1m + v0\right) \tag{3}$$

2.7 *Confusion Matrix*

Confusion matrix adalah suatu metode yang biasanya digunakan untuk melakukan perhitungan akurasi. *Confusion Matrix* dapat digunakan untuk menilai seberapa baik contoh pengelompokan dapat mengenali data dari berbagai tingkat. Tingkat akurasi model klasifikasi adalah indikator kinerjanya [16]. Selain untuk menghitung seberapa benar akurasi sebuah metode, confusion matrix dapat digunakan untuk menghitung nilai *recall, precision, dan F-1 Score* [19]. Isi dari *confusion matrix* bisa dibilang sama dengan *matrix* bujur sangkar (*matrix* yang memiliki dimensi sama) dengan berukuran N x N, di mana N adalah jumlah kelas yang tersedia dalam klasifikasi data. Matrix ini memiliki baris dan kolom yang menunjukkan total kelas data sebenarnya. Struktur dari *confusion matrix* dapat dilihat dari Gambar 2



Gambar 2 Struktur Tabel *Confusion Matrix*

Berdasarkan struktur tabel *confusion matrix* di atas, *confusion matrix* terdiri dari empat nilai;

1. TP (nilai benar positif) adalah jumlah prediksi yang benar di mana model memprediksi kebenaran positif dari suatu data, dan ternyata data tersebut benar positif;
2. TN (nilai benar negatif) adalah jumlah prediksi yang benar di mana model memprediksi kebenaran negatif dari suatu data, dan ternyata data tersebut benar negatif.
3. FP adalah jumlah prediksi yang salah, di mana model memprediksi kebenaran positif dari suatu data, tetapi sebenarnya data itu negatif.
4. FN adalah jumlah prediksi yang salah, di mana model memprediksi kebenaran negatif dari suatu data, tetapi sebenarnya data itu positif [20].

ISSN 1410-7228 (Versi Cetak)
ISSN 2716-4233 (Versi Elektronik)

Evaluasi *Accuracy, Precision, Recall, F1-Score* tersebut diperoleh dengan menggunakan Persamaan (4), (5), (6), dan (7) [18]:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

$$Precision = \frac{TP}{TP + FP} \tag{5}$$

$$Recall = \frac{TP}{TP + FN} \tag{6}$$

$$F1 - Score = \frac{2 * Recall * Precision}{Recall + Precision} \tag{7}$$

3. HASIL DAN PEMBAHASAN

Hasil exprimen dari dataset yang digunakan dalam penelitian ini ditunjukkan di bagian ini. Dataset yang digunakan adalah Indeks Standar Pencemaran Udara Jakarta; sumbernya dijelaskan dalam metode penelitian tentang pengumpulan data. yang digunakan dalam penelitian ini adalah data dari tahun 2014 – 2019. Dan setelah digabungkan bertotal 6556 data dengan 8 fitur, data ini masih dalam bentuk data kotor (belum melakukan *data cleansing), K-Nearest Neighbor* dan *Artificial Neural Network adalah* dua algoritma yang digunakan untuk melakukan klasifikasi pada dataset ini. Keduanya diimplementasikan menggunakan *Python*, bahasa pemograman. Hasil penelitian ini adalah untuk membandingkan dua algoritma untuk menentukan algoritma mana yang paling cocok untuk dataset Indeks Standar Pencemaran Udara wilayah Jakarta saat ini.

Berikut adalah Penjelasan dari fitur fitur yang ada di dataset ini:

1. **Periode** : tahun dan bulan yang digunakan untuk mengukur data
2. **Tanggal** : tanggal pengukuran kualitas udara
3. **Stasiun** : lokasi pengukuran kualitas data.
4. **PM 10**: Pengukuran partikulat
5. **SO 2**: Pengukuran sulfida (dalam bentuk SO2)
6. **CO**: Pengukuran karbon monoksida
7. **O3**: Pengukuran ozon 7. NO2: Pengukuran nitrogen dioksida
8. **Max**: Pengukuran nilai paling tinggi pada saat yang sama
9. **Critical**:” adalah parameter dengan hasil pengukuran paling tinggi
10 **Kategori**: Kelas target dengan nilai Sedang, Baik, Tidak Sehat, dan Sangat Tidak Sehat

Sebelum melakukan pembuatan model untuk klasifikasi, tahap *data cleansing* akan dilakukan terlebih dahulu. Tahap ini dilakukan untuk menghilangkan data data yang memiliki nilai null dan data duplikat pada dataset ini. Setelah itu fitur fitur yang tidak berpengaruh akan dilakukan penghilangan untuk melakukan prediksi yang dibuat oleh model. Pada dataset ini, fitur “Periode”, “Tanggal”, dan “Stasiun” tidak relevan dan tidak terlalu memengaruhi hasil prediksi yang dibuat oleh model. Fitur “Critical” juga tidak terlalu memengaruhi hasil prediksi yang dibuat oleh model karena sudah ada di fitur yang sama yang Bernama “Max”. oleh karena itu fitur tersebut akan di hilangkan dari dataset sebelum kita melakukan pembuatan dan pelatihan model.

Setelah langkah *data cleansing*, langkah eksperimen akan dilanjutkan dengan tahap *splitting data*. tahap ini akan membagi *dataset* menjadi dua, yaitu data latih dan data uji. Presentase data latih dan data uji yang dipakai adalah 80% data latih dan 20% data uji.

Dengan selesainya tahap *splitting data*, klasifikasi data akan dilakukan. Untuk eksperimen pertama metode klasifikasi ANN akan digunakan. ANN ini dijalankan dengan memasukan jumlah iterasi terlebih dahulu. Jumlah iterasi yang digunakan pada eksperimen ini adalah 100, 250, 400, 550, dan 700. Hasil akurasi dari masing masing jumlah iterasi dapat dilihat pada Tabel 1

Tabel 1. Akurasi ANN berdasarkan Jumlah Iterasi

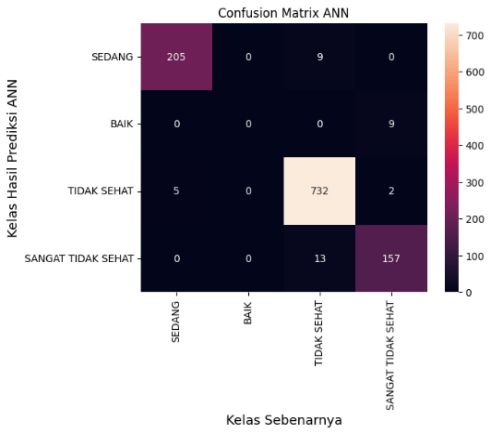
No	Max Iterasi	Accuracy
1	100	0.88
2	250	0.97
3	400	0.97
4	550	0.97
5	700	0.97

Berdasarkan hasil dari Tabel 1, terlihat bahwa untuk maksimum iterasi 100 mendapatkan hasil 88% sedangkan maksimum iterasi 250, 400, 550, dan 700 memiliki akurasi sebesar 97%. Pada eksperimen ini maksimum iterasi yang digunakan adalah 400, karena memiliki nilai akurasi sebesar 97%. Hasil metrik evaluasi seperti *Accuracy, Precision, Recall, dan F-1 Score* untuk maksimum iterasi 400 dapat dilihat pada tabel 2. Dan hasil dari *confusion matrix* dari algoritma ini bisa dilihat pada Gambar 3.

Tabel 2. Hasil metrik evaluasi ANN dengan max_iter = 400

Accuracy	Precision	Recall	F-1 Score
0.97	0.96	0.97	

Pada Gambar 3 dapat dilihat dari 214 data uji dengan hasil “SEDANG”, model ANN dapat memprediksi 205 data dengan tepat, dan hanya 9 data yang salah prediksi. data uji untuk hasil “BAIK” adalah 9, algoritma ANN ini salah memprediksi dan menganggap data tersebut adalah data “SANGAT TIDAK SEHAT”. Data uji untuk hasil “TIDAK SEHAT” adalah 739 data, dari 739 data tersebut model ANN dapat memprediksi 732 data secara benar dan hanya 7 data yang salah di prediksi, dari 7 data tersebut model ANN memasukan 5 data ke dalam hasil “SEDANG” dan 2 data dimasukan ke “SANGAT TIDAK SEHAT”. dan untuk data uji hasil “SANGAT TIDAK SEHAT” memiliki total 170 data uji, model ANN memprediksi 157 data dengan benar dan 13 data salah.



Gambar 3. *Confusion matrix* untuk algoritma ANN dengan max_iter = 400

Pembuatan model untuk Algoritma KNN dilakukan dengan memasukan variasi nilai K (Jumlah Tetangga). Dalam eksperiment ini nilai K yang digunakan adalah 2,3,5,8,12, dan 17. Dari nilai K tersebut, nilai K yang akan digunakan adalah nilai K yang memiliki akurasi tertinggi. Hasil akurasi dari masing masing nilai K dapat dilihat pada tabel 3.

ISSN 1410-7228 (Versi Cetak)
ISSN 2716-4233 (Versi Elektronik)

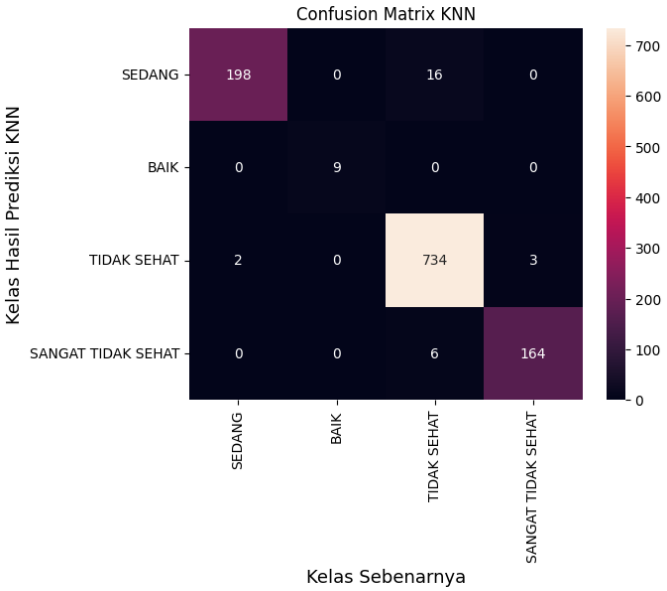
Tabel 3. Akurasi KNN berdasarkan Nilai K

K	Accuracy
2	0.97
3	0.97
5	0.98
8	0.98
12	0.98
17	0.98

Berdasarkan Hasil Akurasi pada Tabel 3 diatas, untuk nilai K 2 dan 3 menghasilkan akurasi 97% untuk nilai K dari 5, 8, 12, dan 17 menghasilkan akurasi 98 %. Dikarenakan ada peningkatan akurasi pada saat nilai K = 5 dan tidak ada perubahan akurasi sampai nilai K = 17, karena itulah nilai K yang akan digunakan untuk perbandingan dengan algoritma ANN adalah K = 5. Hasil metrik evaluasi KNN dengan nilai K = 5 dapat dilihat pada Tabel 4, dan hasil *confusion matrix* dapat dilihat ada Gambar 4.

Tabel 4. Metrik evaluasi KNN dengan nilai k = 5

	Accuracy	Precision	Recall	F-1 Score
KNN	0.98	0.99	0.97	0.98



Gambar 4. *Confusion matrix* untuk algoritma KNN dengan nilai K = 5

Pada Gambar 4 diatas dapat dilihat dari 214 data uji dengan hasil “SEDANG”, model KNN dapat memprediksi 198 data dengan tepat, menyisakan 16 data yang salah di prediksi. data uji untuk hasil “BAIK” adalah 9, algoritma KNN ini memprediksi data uji tersebut dengan benar. Data uji untuk hasil “TIDAK SEHAT” adalah 739 data, dari 739 data tersebut model KNN dapat memprediksi 734 data secara benar dan hanya 5 data yang salah di prediksi, dari 5 data tersebut model KNN memasukan 2 data ke dalam hasil “SEDANG” dan 3 data dimasukan ke “SANGAT TIDAK SEHAT”. dan untuk data uji yang terakhir dengan hasil “SANGAT TIDAK SEHAT” memiliki total 170 data uji, model KNN memprediksi 164 data dengan benar dan 6 data salah.
Untuk Perbandingan hasil metrik evaluasi antara algotima Artificial Neural Network (ANN) dan K-Nearest Neigbor (KNN) dapat dilihat pada Tabel 5.

Tabel 4. Perbandingan Metrik evaluasi KNN dan ANN

	Accuracy	Precision	Recall	F-1 Score
ANN	0.97	0.96	0.97	0.96
KNN	0.98	0.99	0.97	0.98

4. KESIMPULAN

Dalam Penelitian ini, hasil dari eksperimen membandingkan kedua algoritma *machine learning* yaitu algoritma *K-Nearest Neighbor* (KNN) dan algoritma *Artificial Neural Network* (ANN), dalam memprediksi data Indeks Standar Pencemaran Udara (ISPU) di wilayah Jakarta. Hasil eksperimen ini menunjukan bahwa algoritma KNN dapat mencapai akurasi 98%, sedangkan algoritma ANN memiliki akurasi 97%. Dari hasil perbandingan tersebut dapat kita simpulkan bahwa algoritma KNN adalah algoritma yang lebih cocok dibandingkan dengan algoritma ANN dalam memprediksi dataset pada penelitian ini.

UCAPAN TERIMA KASIH

Penulis ingin menyampaikan rasa terima kasih kepada Tuhan Yang Maha Esa, Bu Teny Handayani sebagai Dosen Mata Kuliah Machine Learning, serta kepada teman-teman dan semua pihak yang telah memberikan bantuan dan dukungan yang sangat berarti selama proses penulisan jurnal ini. Saya juga ingin mengucapkan terima kasih kepada Jurnal Computatio dan IJCCS sebagai landasan dalam format makalah ini

DAFTAR PUSTAKA

[1] F. M. Putra and I. S. Sitanggang, "Classification model of air quality in Jakarta using decision tree algorithm based on air pollutant standard index," *IOP Conference Series: Earth and Environmental Science*, pp. 1-11, 2020.

[2] M. A. Latief and Y. Karyanti, "DATA MINING & ANALYTIC FORECASTING INDEKS STANDAR PENCEMAR UDARA JAKARTA MENGGUNAKAN METODE LINEAR REGRESSION (STUDI KASUS: DATASET INDEKS STANDAR PENCEMAR UDARA JAKARTA 2021)," *Journal of Social Resarch*, p. 11, 20 September 2022.

[3] Sang, E. Sutoyo and I. Darmawan, "ANALISIS DATA MINING UNTUK KLASIFIKASI DATA KUALITAS UDARA," *e-Proceeding of Engineering*, vol. 8, no. 5, p. 10, Oktober 2021.

[4] W. Apriliah, I. Kurniawan, M. Baydhowi and T. Haryati, "Prediksi Kemungkinan Diabetes pada Tahap Awal Menggunakan Algoritma KlasifikasiRandom Forest," *SISTEMASI: Jurnal Sistem Informas*, vol. 10, no. 1, pp. 1-9, Januari 2021

[5] P. R. Sihombing and A. M. Arsani, "COMPARISON OF MACHINE LEARNING METHODS IN CLASSIFYING POVERTY IN INDONESIA IN 2018," *Jurnal Teknik Informatika (JUTIF)*, vol. 2, no. 1, p. 6, 15 Januari 2021.

[6] J. Narabel and S. Budi, "Deteksi Dini Status Keanggotaan Industri Kebugaran Menggunakan Pendekatan Supervised Learning," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 6, no. 2, p. 12, 2020

[7] S. Nurjanah, A. M. S. Universitas and D. S. Kusumaningrum, "PENERAPAN ALGORITMA K – NEAREST NEIGHBOR (KNN) UNTUK KLASIFIKASI PENCEMARAN UDARA DI KOTA JAKARTA," *Scientific Student Journal for Information, Technology and*, vol. 1, no. 2, p. 6, July 2020.

[8] L. A. Putri and Suwanda, "Implementasi Metode Artificial Neural Network (ANN) Algoritma Backpropagation untuk Klasifikasi Kualitas Udara di Provinsi DKI Jakarta Tahun 2021," *Bandung Conference Series: Statistics*, 2022.

[9] D. Cahyanti and A. R. S. A. Husniar, "Analisis performa metode Knn pada Dataset pasien pengidap Kanker Payudara," *Indonesian Journal of Data and Science*, vol. 1, no. 2, p. 5, July 2020.

ISSN 1410-7228 (Versi Cetak)
ISSN 2716-4233 (Versi Elektronik)

[10] M. Resha, S. Syamsu, Andryanto and Bakry, "Prediksi Penyebaran Kasus Tuberkulosis dengan metode Artificial Neural Network dan Multi-Layer Perceptron di kota makassar," *JNSTA JOURNAL OF NATURAL SCIENCE AND TECHNOLOGY ADPERTISI,* p. 6, Desember 2021.

[11] A. Amalia, A. Zaidah and I. N. Isnainiyah, "PREDIKSI KUALITAS UDARA MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOR," *Jurnal Ilmiah Penelitian dan Pembelajaran Informatika,* vol. 7, no. 2, p. 12, Juni 2022

[12] J. E. Widyaya and S. Budi, "Pengaruh Preprocessing Terhadap Klasifikasi Diabetic Retinopathy dengan Pendekatan Transfer Learning Convolutional Neural Network," *Jurnal Teknik Informatika dan Sistem Informasi,* vol. 7, no. 1, p. 15, 4 Maret 2021

[13] A. Putri, C. S. Hardiana, E. Novfuja, F. T. P. Siregar, Rahmaddeni, Y. Fatma and R. Wahyuni, "Komparasi Algoritma K-NN, Naive Bayes dan SVM untuk Prediksi Kelulusan Mahasiswa TingkatAkhir," *MALCOM: Indonesian Journal of Machine Learning and Computer Science,* vol. 3, no. 1, p. 7, April 2023.

[14] S. D. Prasetyo, S. ShofiahHilabi and F. Nurapriani, "Analisis Sentimen Relokasi Ibukota Nusantara Menggunakan Algoritma Naïve Bayes dan KNN," *Jurnal KomtekInfo,* vol. 10, no. 1, p. 7, 2023

[15] A. R. Bohari, "Meningkatkan Kinerja Backpropagation Neural Network Menggunakan Algoritma Adaptif," *IMTechno: Journal of Industrial Management and Technology,* vol. 3, no. 1, p. 6, Januari 2022

[16] N. R. Sari and Y. Mar'atullatifah, "PENERAPAN MULTILAYER PERCEPTRON UNTUK IDENTIFIKASI KANKER PAYUDARA," *Jurnal Cakrawala Ilmiah,* vol. 2, no. 8, p. 8, April 2023.

[17] D. Fatmawati, W. Trisnawati, Y. Jumaryadi and G. Triyono, "Klasifikasi Tingkat Kepuasan Penggunaan Layanan Teknologi Informasi Menggunakan Decision Tree," *KLIK: Kajian Ilmiah Informatika dan Komputer,* vol. 3, no. 6, p. 7, Juni 2023.

[18] M. L. B. Permadi and R. Gumilang, "PENERAPAN ALGORITMA CNN (CONVOLUTIONAL NEURAL NETWORK) UNTUK DETEKSI DAN KLASIFIKASI TARGET MILITER BERDASARKAN CITRA SATELIT," *Jurnal Sosial dan Teknologi (SOSTECH),* vol. 4, no. 2, p. 10, Febuari 2024.

[19] N. Puspitasari, A. Septiarini and A. R. Aliudin, "METODE K-NEAREST NEIGHBORDAN FITURWARNAUNTUK KLASIFIKASI DAUN SIRIH BERDASARKAN CITRA DIGITAL," *Jurnal PROSISKO,* vol. 10, no. 2, p. 8, 2023.

[20] A. Nugroho and Y. Religia, "Analisis Optimasi Algoritma Klasifikasi Naive Bayes menggunakan Genetic Algorithm dan Bagging," *JURNAL RESTI,* vol. 5, no. 3, pp. 504-510, 2021.