

Implementasi Algoritma Modified K-Nearest Neighbor untuk Klasifikasi Indeks Kualitas Udara Perkotaan di Berbagai Negara

Ni Made Ayu Pranasanthi Dewi^{a1}, I Made Widiartha^{a2}

Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Udayana
Jalan Raya Kampus UNUD, Bukit Jimbaran, Kuta Selatan, Badung, Bali, Indonesia
¹gekana744@gmail.com
²madewidiartha@unud.ac.id

Abstract

One important component for the sustainability of living organisms is the availability of clean air. Air pollution has detrimental repercussions such as deteriorating air quality, which can have harmful impacts on the environment. As a result, modeling and monitoring air quality are essential steps in reducing air pollution, and the Air Quality Index allows for the review of this monitoring. This study's objective is to apply the Modified K-Nearest Neighbor method in classifying urban Air Quality Index data from various countries. The "World Air Quality Index by City and Coordinates" Kaggle website provided the data that were used. The data were then processed and tested using the Modified K-Nearest Neighbor model, this builds on the K-Nearest Neighbor technique. 16,695 data in total were divided into 80% training data (13,356 data) and 20% testing data (3,339 data) in order to conduct testing. Evaluation was performed by comparing the performance of the Modified K-Nearest Neighbor method with the traditional K-Nearest Neighbor method. The Modified K-Nearest Neighbor approach with $K=1$ produced an accuracy rate of 99.73% during testing, whereas the K-Nearest Neighbor method produced an accuracy rate of 99.64%. Using the K-Fold Cross Validation, K-Nearest Neighbor method perform highest mean score of 99,04% and Modified K-Nearest Neighbor perform highest mean score of 99,46%.

Keywords: Air Pollution, Air Quality Index, Classification, K-Nearest Neighbor, Modified K-Nearest Neighbor.

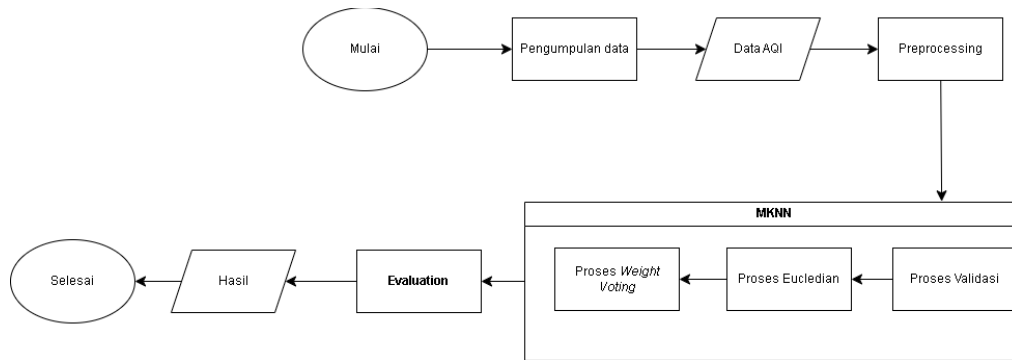
1. Pendahuluan

Salah satu komponen yang memiliki peran penting dalam kelangsungan makhluk hidup adalah ketersediaan udara. Adanya kontaminan yang terdapat dalam udara memberikan efek negatif berupa penurunan kualitas udara. Kualitas udara atau *air quality* merupakan kadar kandungan udara berdasarkan konsentrasi polutan di lokasi tertentu [1], semakin tinggi kontaminan yang terdapat di suatu daerah maka berbanding lurus dengan tingkat polusi udara di daerah tersebut. Polusi udara dapat berasal dari berbagai aktivitas seperti bencana alam, emisi industri, pembakaran, serta aktivitas manusia lainnya. Komponen utama polutan udara meliputi partikel halus ($PM_{2.5}$), Nitrogen Dioksida (NO_2), Sulfur Dioksida (SO_2), Karbon Monoksida (CO), dan Ozon (O_3) [2]. Menurut WHO, dampak berbahaya dari polusi udara menyebabkan sekitar 7 juta kematian setiap tahunnya dan berkontribusi pada berbagai masalah kesehatan seperti penyakit pernapasan, penyakit kardiovaskular, dan kanker [3]. Kabut asap yang disebabkan oleh kebakaran hutan dan lahan atau karhutla mengakibatkan setidaknya 6.025 warga menderita infeksi saluran pernapasan akut (ISPA) [4]. Oleh karena itu pemantauan dan pemodelan kualitas udara menjadi sangat penting dalam upaya untuk menanggulangi polusi udara, pemantauan ini dapat ditinjau dari Indeks Kualitas Udara. *Air Quality Index* (AQI) atau Indeks Kualitas Udara adalah suatu sistem yang digunakan untuk mengukur, menilai, dan melaporkan kualitas udara pada suatu wilayah. Meninjau uraian tersebut maka sangat penting dilakukan deteksi indeks kualitas udara secara berkala sehingga dapat dilakukan pengontrolan yang sesuai untuk menghindari risiko kerusakan lingkungan lebih lanjut yang dapat memengaruhi kesehatan

masyarakat. Dalam menyelesaikan permasalahan ini, salah satu pendekatan yang dapat digunakan adalah dengan mengaplikasikan metode klasifikasi menggunakan model *Machine Learning*. *Machine Learning (ML)* atau Pembelajaran Mesin adalah cabang dari *Artificial Intelligence (AI)* yang fokus belajar dari data (*learn from data*), yaitu fokus pada pengembangan sistem yang mampu belajar secara "mandiri" tanpa harus berulang kali diprogram manusia [5]. Metode klasifikasi menggunakan *ML* dapat dilakukan dengan beberapa algoritma seperti *K-Nearest Neighbor (KNN)*, *Naïve Bayes*, dan *Support Vector Machine (SVM)*. Berdasarkan penelitian sebelumnya yang berjudul "*Comparative Analysis of Restock Needs Bottled Water using K-Nearest Neighbor (K-NN), Support Vector Machine (SVM), and the Naïve Bayes Algorithm*" mengenai analisis perbandingan kebutuhan pengisian ulang air minum botol dengan dengan algoritma *K-Nearest Neighbor*, *Support Vector Machine*, dan *Naïve Bayes* menunjukkan akurasi bahwa metode klasifikasi menggunakan algoritma *K-Nearest Neighbor* mampu memberikan hasil terbaik dengan nilai akurasi 88,20% dibandingkan dengan algoritma *Naïve Bayes* (66,20%) dan *Support Vector Machine* (84,51%) [6]. Algoritma *K-Nearest Neighbor* merupakan metode yang menerapkan prinsip "*nearest neighbor*" dimana prediksi untuk sebuah instansi baru didasarkan pada label atau nilai-nilai yang dimiliki oleh tetangga terdekatnya dalam ruang fitur yang ditentukan dari sejumlah *K* tetangga terdekat yang akan dipertimbangkan untuk melakukan prediksi [3]. Algoritma *K-Nearest Neighbor* memiliki kelemahan, salah satunya yaitu kelas objek baru ditentukan berdasarkan *voting* mayoritas kelas pada *K* jarak terdekat [7]. Maka dari itu dilakukan beberapa modifikasi untuk mengembangkan performa algoritma ini, yaitu dengan penambahan proses klasifikasi baru yang mencakup perhitungan nilai validitas untuk menilai validitas data latih dan perhitungan suara berat untuk menentukan bobot masing-masing terdekat. Pengembangan algoritma *K-Nearest Neighbor* ini selanjutnya disebut dengan algoritma *Modified K-Nearest Neighbor*. Hasil penelitian terdahulu terkait dengan proses klasifikasi tingkat kualitas udara di Provinsi DKI Jakarta menggunakan algoritma *Naïve Bayes* mendapatkan hasil rata rata *accuracy* sebesar 88%, *recall* 96%, *precision* 85%, dan *f1-score* 90% [8]. Penelitian topik serupa dengan menerapkan algoritma *Random Forest Classifier* memperoleh hasil terbaik pada akurasi 90% [9]. Penelitian serupa berjudul "Klasifikasi *Data Mining* untuk Menentukan Kualitas Udara di Provinsi DKI Jakarta Menggunakan Algoritma *K-Nearest Neighbors (K-NN)*" memperoleh hasil akurasi menggunakan *K-5fold* sebesar 75,35% [3]. Pada penelitian berjudul "*Feature Selection using Information Gain on the K-Nearest Neighbor (KNN) and Modified K-Nearest Neighbor (MKNN) Methods for Chronic Kidney Disease Classification*" mengenai klasifikasi data penyakit ginjal kronis yang membandingkan algoritma *Modified K-Nearest Neighbor (MKNN)* dan *K-Nearest Neighbor (KNN)* dengan 400 data yang terbagi menjadi 2 kelas diperoleh hasil akurasi tertinggi *MKNN* sebesar 98%, lebih tinggi 1,39% dibanding *KNN* dengan akurasi tertinggi 96,61% [10]. Kemudian penelitian selanjutnya dengan judul "Implementasi Algoritma *Modified K-Nearest Neighbor (MKNN)* untuk Klasifikasi Penyakit Kanker Payudara memperoleh akurasi dengan *K=1* sebesar 97,61% [11]. Berdasarkan uraian diatas, penelitian ini mencoba untuk memberikan penyelesaian klasifikasi indeks kualitas udara dengan penerapan algoritma *Modified K-Nearest Neighbor (MKNN)* sebagai pengoptimasian dari penelitian sebelumnya. Penelitian ini menerapkan beberapa tahapan baru dalam algoritma *K-Nearest Neighbor (KNN)* sebagai modifikasi untuk dapat memberikan hasil akurasi yang lebih tinggi dan menggunakan pengujian variasi nilai *K* serta *K-fold Cross Validation*. Tujuan penelitian ini yaitu untuk mengimplementasikan dan menganalisis algoritma *MKNN* dalam mengklasifikasikan indeks kualitas udara pada sejumlah 174 negara di seluruh dunia.

2. Metode Penelitian

Penelitian ini mengimplementasikan algoritma *Modified K-Nearest Neighbor* untuk proses klasifikasi data indeks kualitas udara menggunakan *input* berupa data *Air Quality Index (AQI)* dari 174 negara. *Output* yang akan dihasilkan adalah hasil evaluasi dari algoritma *Modified K-Nearest Neighbor* dan *K-Nearest Neighbor* tradisional. Alur penelitian yang digunakan terdiri dari 5 tahapan dapat dilihat pada gambar 1, dimulai dari proses pengumpulan data, melakukan *preprocessing*, membagi data, melakukan klasifikasi dengan algoritma *Modified K-Nearest Neighbor* untuk menghasilkan hasil evaluasi berupa nilai akurasi.



Gambar 1. Alur penelitian

2.1. Pengumpulan Data

Data yang diaplikasikan dalam penelitian ini berupa data sekunder yaitu dataset *AQI and Lat Long of Countries* yang bersumber dari situs *Kaggle* yang berjumlah 16695 data. Dataset ini terakhir diperbarui pada tahun 2023. Dalam dataset ini terdapat informasi mengenai tingkat polusi udara di berbagai negara di dunia yang ditinjau dari *Air Quality Index* masing-masing negara. Karakteristik data ini terdiri dari 14 kolom dengan atribut berupa *country*, *city*, *AQI category*, *AQI value*, *AQI CO category*, *AQI CO value*, *AQI NO2 category*, *AQI NO2 value*, *AQI Ozone value*, *AQI Ozone category*, *AQI PM2.5 value*, *AQI PM2.5 category*, *latitude*, serta *logitude* dengan format file berupa *csv*. Data ini menampilkan indeks kualitas udara dari sejumlah 174 negara dengan tampilan 5 data teratas pada gambar 2. Fitur yang digunakan dalam penelitian ini berupa *AQI value*, *AQI CO category*, *AQI CO value*, *AQI NO2 category*, *AQI NO2 value*, *AQI Ozone value*, *AQI Ozone category*, *AQI PM2.5 value*, *AQI PM2.5 category* dan terdapat 5 kelas *output* dalam atribut *AQI category* yang digunakan sebagai label yaitu *good*, *moderate*, *unhealthy for sensitive group*, *unhealthy*, *very unhealthy dan hazardous*. Pada tabel 1 merupakan tabel atribut dan domain yang digunakan dalam penelitian ini.

Gambar 2. Atribut dataset *Air Quality Index*

Country	City	AQI Value	AQI Category	CO AQI Value	CO AQI Category	Ozone AQI Value	Ozone AQI Category	NO2 AQI Value	NO2 AQI Category	PM2.5 AQI Value	PM2.5 AQI Category	lat	Ing
Russian Federation	Praskoveya	51	Moderate	1	Good	36	Good	0	Good	51	Moderate	44.7444	44.2031
Brazil	Presidente Dutra	41	Good	1	Good	5	Good	1	Good	41	Good	-5.2900	-44.4900
Brazil	Presidente Dutra	41	Good	1	Good	5	Good	1	Good	41	Good	-11.2958	-41.9869
Italy	Priolo Gargallo	66	Moderate	1	Good	39	Good	2	Good	66	Moderate	37.1667	15.1833
Poland	Przasnysz	34	Good	1	Good	34	Good	0	Good	20	Good	53.0167	20.8833

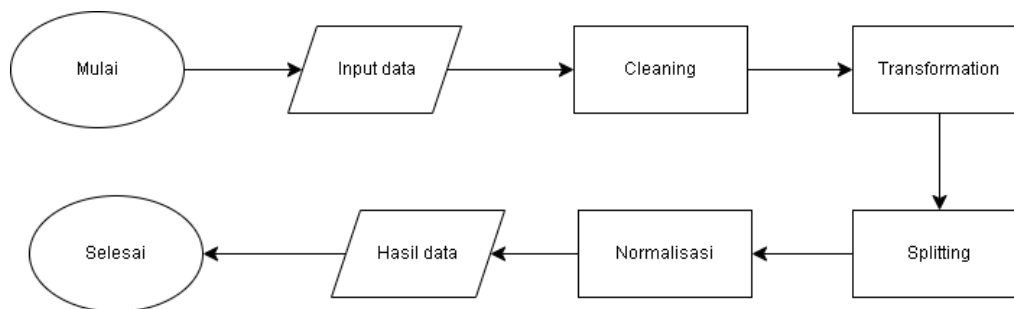
Tabel 1. Atribut dan domain dalam penelitian

Atribut	Domain
AQI Value	9 - 444.
AQI Category	<i>Moderate, Good, Unhealthy, Very Unhealthy, Unhealthy for Sensitive Groups, Hazardous.</i>
CO AQI Value	0 -133.
CO AQI Category	<i>Good, Unhealthy for Sensitive Groups, Moderate.</i>
Ozone AQI Value	0 – 210.

Atribut	Domain
Ozone AQI Category	<i>Moderate, Good, Unhealthy, Unhealthy for Sensitive Groups, Very Unhealthy.</i>
NO2 AQI Value	0 – 91.
NO2 AQI Category	<i>Good, Moderate.</i>
PM2.5 AQI Value	3 – 475.
PM2.5 AQI Category	<i>Moderate, Good, Unhealthy, Very Unhealthy, Unhealthy for Sensitive Groups, Hazardous.</i>

2.2. Preprocessing

Preprocessing adalah tahapan awal untuk menyiapkan data indeks kualitas udara yang telah didapat menjadi data yang siap diolah ke tahap klasifikasi menggunakan algoritma *Modified K-Nearest Neighbor*. Berikut merupakan alur *preprocessing* yang diterapkan dalam penelitian ini terdapat dalam gambar 3.



Gambar 3. Alur *Preprocessing*

2.2.1. Cleaning

Tahap ini merupakan tahap pembersihan data dari data ganda, *missing value*, serta data-data yang tidak relevan untuk mendapat hasil data yang sesuai. Adapun atribut yang tidak diperlukan dalam penelitian ini yaitu garis lintang (*lng*) dan garis bujur (*lat*) negara negara, serta nama negara dan kota.

2.2.2. Transformation

Tahap ini merupakan tahap memperbaiki atribut yang belum sesuai kategori agar sesuai dengan kategori yang dibutuhkan dalam algoritma *Modified K-Nearest Neighbor*. Pada tahap ini dijalankan metode *Label Encoder* yang dapat mentransformasi nilai atribut yang masih berupa *string* menjadi *integer* agar dapat diklasifikasi.

2.2.3. Splitting

Pada langkah ini, data dibagi menjadi dua bagian yaitu data *training* untuk melatih algoritma dan data *testing* untuk menguji. Sejumlah 16695 data akan dibagi dengan persentase data training sebesar 80%(13356 data) serta data testing sebesar 20%(3339 data).

2.2.4. Normalisasi

Tahap ini dijalankan menggunakan metode *Standard Scaler* yang memproses nilai-nilai dalam dataset untuk ditransformasi sehingga menghasilkan nilai rata-rata yaitu 0 serta standar deviasi yaitu 1. Hal ini dilakukan sebagai upaya untuk menjaga konsistensi dan keseimbangan fitur-fitur yang digunakan agar dapat mencapai nilai akurasi yang maksimal.

2.3. Klasifikasi *K-Nearest Neighbor* (KNN)

Dalam *data mining*, algoritma *K-Nearest Neighbor* mengklasifikasikan data berdasarkan jarak terdekat objek. Prinsip kerja algoritma ini berfokus pada pencarian jarak terdekat anatar data *testing* dengan *K* tetangga terdekatnya pada data *training*. Untuk menentukan jarak antara data ini digunakan proses *Euclidian* dengan metode tercantum dalam persamaan (1). Proses ini diterapkan pada data *training* (*x*) dan data *testing* (*y*).

$$d(x_i, y_i) = \sqrt{\sum_{i=0}^n (x_i - y_i)^2} \tag{1}$$

Keterangan:

- d : Jarak *Euclidian*
- i : nilai atribut
- n : dimensi atribut

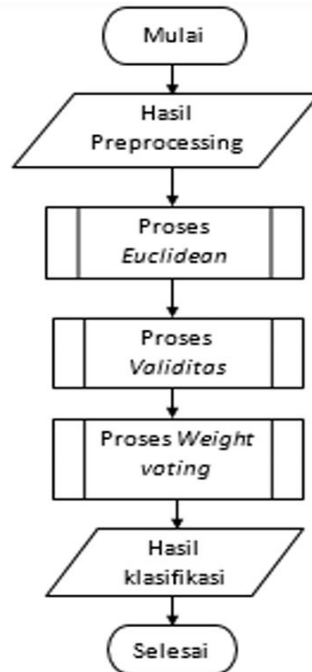
Tahapan proses dalam metode *K-Nearest Neighbor* digambarkan dalam diagram alur pada gambar 4.



Gambar 4. Proses MKNN

2.4. Klasifikasi *Modified K-Nearest Neighbor* (MKNN)

Modified K-Nearest Neighbor (MKNN) merupakan metode yang digunakan untuk mengidentifikasi label kelas data baru yang sesuai dengan *k* validasi data yang telah ditetapkan melalui perhitungan *K-Nearest Neighbor* (KNN). Tujuan utama modifikasi metode ini adalah untuk menentukan kelas label dari query instance ke dalam *K* data latih yang telah divalidasi. Terdapat 2 proses tambahan yaitu proses validasi dan proses *weight voting* dengan alur proses ditunjukkan pada gambar 5.



Gambar 5. Proses MKNN

2.4.1. Proses Validasi

Proses ini merupakan tahapan menghitung validitas data *training* yang akan digunakan. Nilai validitas data tergantung pada setiap tetangganya dan diterapkan bersamaan untuk seluruh data *training* secara serentak. Antara H tetangga terdekat dari data *training* (x), ditentukan berapa banyak data yang memiliki label yang sama dengan x . Adapun persamaan untuk menghitung validitas tiap data *training* ditunjukkan pada persamaan (2).

$$Validity(x) = \frac{1}{K} \sum_{i=1}^K S(label(x))(N_i(x)) \quad (2)$$

Keterangan :

- K : Jumlah titik terdekat
- $lbl(x)$: kelas x
- $lbl(N_i)$: label kelas titik terdekat

Dalam proses ini, persamaan S pada (3) digunakan untuk menentukan kesamaan antara titik x dan data ke- i dari tetangga terdekat. Persamaan ini menunjukkan bahwa label kelas kategori a bernilai 1 jika label kelas kategori b sama dengan label kelas a , dan bernilai 0 jika label kelas kategori a tidak sama dengan label kelas b .

$$S(a, b) = \{1 \ a = b, 0 \ a \neq b\} \quad (3)$$

Keterangan :

- A : Kelas a pada data training
- b : Kelas selain kelas a pada data training

2.4.2. Proses Weight Voting

Dari hasil perhitungan validitas menggunakan persamaan (2) dan (3) pada masing-masing K tetangga terdekat serta menghitung jarak *Euclidean* dengan persamaan (1), selanjutnya dalam persamaan (4) akan dilakukan perhitungan *weight voting*.

$$W(i) = \text{Validity}(i) \times \frac{1}{de(i)+0.5} \quad (4)$$

Keterangan:

$W(x)$: Nilai *weight voting*
 $d(x,y)$: Jarak *Euclidian*
0,5 : Konstanta

2.5. Pengujian

Pengujian dilakukan untuk mengetahui kinerja dari model yang digunakan dalam mengklasifikasi indeks kualitas udara. Dalam pengujian algoritma *Modified K-Nearest Neighbor* menyertakan *K-Nearest Neighbor* sebagai pembandingan, diperlukan *value K* yang tepat agar menghasilkan akurasi maksimal. Penentuan nilai *K* dilakukan dengan melakukan pengujian variasi nilai *K* berupa dan metode *K-fold Cross Validation*.

2.5.1. Pengujian Variasi Nilai K

Proses pengujian ini menggunakan variasi nilai *K* yang berbeda dengan tujuan untuk mengetahui pengaruh nilai dari tetangga terdekat terhadap nilai akhir akurasi yang diperoleh. Dalam penelitian ini digunakan nilai *K* sebesar $K=1$, $K=3$, $K=5$, $K=7$, dan $K=9$ yang diterapkan pada algoritma *K-Nearest Neighbor* dan *Modified K-Nearest Neighbor*.

2.5.2. Pengujian *K-Fold Cross Validation*

Metode *k-fold* membagi dataset menjadi sejumlah *k*-buah partisi secara acak ke dalam *k* partisi yang berukuran sama dan dilakukan sejumlah *K* kali eksperimen. Tujuan dari pengujian ini adalah untuk mengetahui kelayakan seluruh data dengan sistem yang acak. Pada penelitian ini, pengujian dilakukan dengan membagi dataset menjadi *2 fold*, *3 fold*, *4 fold*, dan *5 fold*. Kemudian hasil nilai *K* tertinggi dari pengujian variasi nilai *K* akan digunakan dalam masing-masing *fold* untuk kedua algoritma yang diuji.

2.6. Evaluasi

Tahap ini dijalankan untuk mengevaluasi performa dari model *Modified K-Nearest Neighbor*. Evaluasi dilakukan dengan melakukan perbandingan nilai akurasi antara model *Modified K-Nearest Neighbor* dan *K-Nearest Neighbor* tradisional.

3. Hasil dan Diskusi

Setelah dilakukan pengujian terhadap model algoritma *K-Nearest Neighbor* dan *Modified K-Nearest Neighbor*, didapat hasil pembandingan kinerja kedua algoritma dalam menjalankan fungsi klasifikasi.

3.1. Pengujian Variasi nilai K

Setelah menjalani tahapan *preprocessing*, data selanjutnya dilatih sekaligus diuji menggunakan pemodelan *K-Nearest Neighbor* dan *Modified K-Nearest Neighbor*. Untuk kedua model digunakan variasi nilai *K* sebesar $K=1$, $K=3$, $K=5$, $K=7$, dan $K=9$. Perbandingan hasil akurasi kedua pemodelan ditunjukkan pada tabel 2. Pada pengujian ini didapat bahwa pemodelan *K-Nearest Neighbor* memiliki nilai akurasi tertinggi pada $K=3$ sebesar 99,64% dan pemodelan *Modified K-Nearest Neighbor* memiliki nilai akurasi tertinggi sebesar 99,73% pada $K=1$.

Tabel 2. Hasil akurasi berdasarkan pengujian variasi nilai K

Nilai K	<i>K-Nearest Neighbor</i>	<i>Modified K-Nearest Neighbor</i>
K=1	99,61%	99,73%
K=3	99,64%	99,67%
K=5	99,55%	99,64%
K=7	99,58%	99,55%
K=9	99,52%	99,64%

3.2. Pengujian K-Fold Cross Validation

Pengujian K-Fold Cross Validation dilakukan dengan menggunakan nilai K terbaik untuk masing-masing pemodelan, yaitu K=3 untuk model *K-Nearest Neighbor* dan K=1 untuk model *Modified K-Nearest Neighbor*. Pada tabel 3 ditunjukkan hasil rata-rata skor K-Fold Cross Validation untuk kedua pemodelan pada kasus fold 2, 3, 4, dan 5. Hasil tersebut menunjukkan bahwa akurasi tertinggi untuk pemodelan *K-Nearest Neighbor* didapat pada pengujian fold=4 sebesar 99,04%. Sedangkan pada pemodelan *Modified K-Nearest Neighbor* didapat hasil akurasi tertinggi pada pengujian fold = 5 sebesar 99,46%.

Tabel 3. Hasil rata rata berdasarkan pengujian variasi nilai K

Nilai fold	<i>K-Nearest Neighbor</i>	<i>Modified K-Nearest Neighbor</i>
2	98,74%	99,13%
3	98,80%	99,25%
4	99,04%	99,52%
5	99,04%	99,46%

4. Kesimpulan

Berdasarkan penelitian yang telah dijalankan, dapat disimpulkan bahwa implementasi pemodelan algoritma *Modified K-Nearest Neighbor* dengan tahapan tambahan berupa validasi dan *weight voting* dapat meningkatkan akurasi sistem klasifikasi dibandingkan dengan menggunakan algoritma *K-Nearest Neighbor* tanpa modifikasi. Namun adanya modifikasi ini tidak menimbulkan peningkatan akurasi yang signifikan. Hal ini dibuktikan dengan nilai akurasi tertinggi pengujian algoritma *Modified K-Nearest Neighbor* lebih tinggi namun hanya sejumlah 0,9% dengan akurasi 99,73% dibandingkan dengan algoritma *K-Nearest Neighbor* dengan akurasi 99,64%. Pada pengujian menggunakan K-Fold Cross Validation didapat rata-rata tertinggi skor pada algoritma *Modified K-Nearest Neighbor* sebesar 99,46% sedangkan pada algoritma *K-Nearest Neighbor* sebesar 99,04% yang menunjukkan peningkatan sebesar 0,42%. Penambahan metode preprocessing untuk penelitian berikutnya diperkirakan mampu untuk menghasilkan peningkatan akurasi yang lebih signifikan.

Daftar Pustaka

- [1] Toha, A. Purwono och W. Gata, "Model Prediksi Kualitas Udara dengan Support Vector Machines dengan Optimasi Hyperparameter GridSearch CV," *Buletin Ilmiah Sarjana Teknik Elektro*, vol. 4, nr 1, pp. 12-21, 2022.
- [2] W. C. Shao och L. Chou, "Political Influence and Air Pollution: Evidence from Chinese Cities," *Heliyon*, vol. 9, nr 7, 2023.
- [3] A. D. Wiranata, Soleman, Irwansyah, I. K. Sudaryana och Rizal, "Klasifikasi Data Mining untuk Menentukan Kualitas Udara di Provinsi DKI Jakarta Menggunakan Algoritma K-

- Nearest Neighbors (K-NN)," *INFOTECH: Journal of Technology Information*, vol. 9, nr 1, pp. 95-101, 2023.
- [4] A. E. Putra och T. Rismawan, "Klasifikasi Kualitas Udara Berdasarkan Indeks Standar Pencemaran Udara (ISPU) Menggunakan Metode Fuzzy Tsukamoto," *Jurnal Komputer dan Aplikasi*, vol. 11, nr 2, pp. 190-196, 2023.
- [5] I. Cholissodin, Sutrisno, A. A. Soebroto, U. Hasanah och Y. I. Febiola, AI, Machine Learning & Deep Learning (Teori & Implementasi), Malang, 2020.
- [6] R. F. D. Pratiwi, S. Sumarlinda och F. E. Nastiti, "Comparative Analysis of Restock Needs Bottled Water using K-Nearest Neighbor (K-NN), Support Vector Machine (SVM), and the Naïve Bayes Algorithm," *International Journal of Information System Technology and Data Science (IJIST-DAS)*, vol. 1, nr 1, pp. 1-8, 2023.
- [7] Y. D. Alfiyanti, D. E. Ratnawati och S. Anam, "Klasifikasi Fungsi Senyawa Aktif Data Berdasarkan Kode Simplified Molecular Input Line Entry System (SMILES) menggunakan Metode Modified K-Nearest Neighbor," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 3, nr 4, pp. 3244-3251, 2019.
- [8] A. A. H. Kirono, I. Asror och Y. F. A. Wibowo, "Klasifikasi Tingkat Kualitas Udara DKI Jakarta Menggunakan Algoritma Naïve Bayes," *e-Proceeding of Engineering*, vol. 9, nr 3, pp. 1962-1969, 2023.
- [9] A. Nugroho, I. Asror och Y. F. A. Wibowo, "Klasifikasi Tingkat Kualitas Udara DKI Jakarta Berdasarkan Open Government Data Menggunakan Algoritma Random Forest," *eProceedings of Engineering*, vol. 10, nr 2, p. 1824–1834, 2023.
- [10] A. Ramadhan, E. Budianita, F. Syafrina och S. Ramadhani, "Feature Selection using Information Gain on the K-Nearest Neighbor (KNN) and Modified K-Nearest Neighbor (MKNN) Methods for Chronic Kidney Disease Classification," Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, 2023.
- [11] M. I. P. Putra, D. T. Murdiansyah och A. Aditsania, "Implementasi Algoritma Modified K-Nearest Neighbor (MKNN) untuk Klasifikasi Penyakit Kanker Payudara," *e-Proceeding of Engineering*, vol. 6, nr 1, pp. 2431-2441, 2019.

Halaman ini sengaja dibiarkan kosong