ANALISIS KLASIFIKASI INDEKS KUALITAS UDARA KOTA DI INDONESIA MENGGUNAKAN METODE K-NEAREST NEIGHBOR DAN NAÏVE BAYES

Ali Maulana [1], Ade Irma Purnamasari [2], Irfan Ali [3]
[1] Teknik Informatika, STMIK IKMI Cirebon
[2] Teknik Informatika, STMIK IKMI Cirebon
[3] Rekayasa Perangkat Lunak, STMIK IKMI Cirebon
Jalan Perjuangan No. 10 B Majasem, Kec. Kesambi, Kota Cirebon, Jawa Barat 45135, Indonesia
alimaulana490@gmail.com

ABSTRAK

Polusi udara adalah masalah yang berdampak buruk pada kehidupan makhluk hidup dan menyebabkan banyak penyakit. Oleh karena itu, penting untuk memantau tingkat pencemaran udara di lingkungan masyarakat. Tujuan dari penelitian ini adalah untuk membandingkan kinerja dari metode *K-Nearest Neighbor* dan *Naïve Bayes*, dalam mengklasifikasikan data Indeks Kualitas Udara kota di Indonesia. Pada penelitian ini, data yang digunakan adalah data *World Air Quality Index by City and Coordinates* yang diperoleh dari situ *Kaggle*. Data ini mencakup atribut-atribut seperti *country, city, AQI value, AQI category* dan lain-lain. Metode penelitian yang digunakan adalah *Knowledge Discovery in Database* (KDD). Selanjutnya, metode *K-Nearest Neighbor* dan *Naïve Bayes* diterapkan pada tahapan data mining menggunakan K-*fold cross validation* dengan percobaan K-2 *fold*, K-3 *fold*, K-4 *fold* dan K-5 *fold*. Evaluasi kinerja akan dilakukan menggunakan metrik-metrik yang relevan seperti akurasi, *precision*, dan *recall*. Berdasarkan hasil klasifikasi didapatkan K-*fold* terbaik yaitu K-5 *fold* dari metode *K-Nearest Neighbor* menghasilkan nilai akurasi 95.13% dan *Naïve Bayes* menghasilkan akurasi 95.97%. Penelitian ini dapat membantu pemerintah dalam pengambilan kebijakan untuk menjaga kualitas udara, dan memberikan informasi kepada masyarakat tentang kualitas udara di lingkungannya.

Kata kunci : *Indeks Kualitas Udara, K-Fold Cross Validation, Klasifikasi, K-Nearest Neighbors, Naïve Bayes*

1. PENDAHULUAN

Polusi udara adalah masalah yang berdampak buruk pada kehidupan makhluk hidup dan menyebabkan banyak penyakit. Oleh karena itu, penting untuk memantau tingkat pencemaran udara di lingkungan masyarakat [1]. Beberapa aktivitas manusia yang dapat mengurangi kualitas udara termasuk asap rokok, aktivitas industri, transportasi, pembakaran hutan dan lahan, dan lainnya [2].

Kualitas udara tidak hanya mempengaruhi ekosistem, tetapi juga kesehatan manusia dan kualitas hidup secara keseluruhan. Pada tahun 2023, Indonesia telah dimasukkan ke dalam daftar negara dengan tingkat polusi tertinggi ke-26 di dunia berdasarkan data *AirVisual* oleh AQI [3]. Dalam hal ini tingkat polusi udara di kota-kota besar di Indonesia telah menjadi masalah yang serius karena dampak negatifnya terhadap kesehatan manusia dan ekosistem.

Pada penelitian sebelumnya yang meneliti tentang kualitas udara dengan menggunakan algoritma *Naïve Bayes* dilakukan oleh Abdul Aziiz Hendri Kirono, dkk menghasilkan nilai rata-rata akurasi 88%, *precision* 85%, *recall* 96%, *f1-score* 90% [4].

Penelitian selanjutnya dilakukan oleh Adityo Nugroho, dkk dengan menggunakan algoritma *Random Forest* menghasilkan kesimpulan klasifikasi data mining dengan algoritma *Random Forest* menghasilkan nilai terbaik pada akurasi 90% [5].

Penelitian selanjutnya dilakukan oleh Devi Dwi Purwanto, dkk dengan menggunakan *Gaussian Naïve Bayes* menghasilkan bahwa klasifikasi dengan metode *gaussian Naïve Bayes*, sistem dapat memberikan akurasi sebesar 93,8% dan memiliki *error rate* sebesar 6,2% [6].

Penelitian selanjutnya yang dilakukan oleh Ade Davy Wiranata, dkk dengan menggunakan algoritma *K-Nearest Neighbors* (K-NN) menghasilkan proses evaluasi algoritma *K-Nearest Neighbors* (K-NN) menggunakan K-5 *fold* diperoleh hasil akurasi K-2 *fold* sebesar 73,97%, akurasi K-3 *fold* sebesar 72,60%, akurasi K-4 *fold* sebesar 72,60%, dan akurasi K-5 *fold* sebesar 75,35% [7].

Dalam penelitian tersebut, hanya ada satu metode yang digunakan untuk mengklasifikasikan data indeks kualitas udara. Oleh karena itu, perbandingan antara metode yang digunakan untuk mengklasifikasikan data indeks kualitas udara menjadi penting dan relevan untuk diteliti. Dengan demikian tujuan utama penelitian ini adalah melakukan perbandingan antara metode *K-Nearest Neighbors* dan *Naïve Bayes* dalam klasifikasi data Indeks Kualitas Udara (IKU) kota di Indonesia dengan membandingkan nilai akurasi, *precision*, dan *recall.*

2. TINJAUAN PUSTAKA

2.1. Algoritma *K-Nearest Neighbor*

Metode *K-Nearest Neighbors* (K-NN) merupakan suatu pendekatan klasifikasi yang menentukan kategori suatu data berdasarkan mayoritas kategori dari k tetangga terdekat. Apabila D menyatakan himpunan data pelatihan, ketika data uji d ditampilkan, algoritma akan mengukur jarak antara setiap data dalam D dan data uji d menggunakan metode jarak Euclidean. Setelah itu, k data dalam D

yang memiliki jarak terdekat dengan d akan dipilih. Kumpulan k data ini membentuk k-nearest neighbors. Akhirnya, kategori dari data uji d ditetapkan berdasarkan mayoritas label kategori pada himpunan k tetangga terdekat tersebut [8]. Secara matematis *Euclidean distance* dapat dituliskan pada persamaan (1) berikut [9]:

$$d(x_i, x_j) = \sqrt{\Sigma_{r=1}^{n} \left(a_r(x_i) - a_r(x_j)\right)^2} \quad (1)$$

Keterangan
$d(x_i, x_j)$: Jarak *Euclidean* (*Euclidean Distance)*
(x_i) : record ke- i
(x_j) : record ke- j
$(a_{r)}$: data ke-r
i,j : 1,2,3,…n

2.2. Algoritma *Naïve Bayes*

Algoritma *Naive Bayes* adalah metode klasifikasi probabilistik yang sederhana dengan menghitung probabilitas dari kumpulan data berdasarkan frekuensi dan kombinasi nilai atribut. Algoritma *Naive Bayes* menggunakan teorema Bayes dengan asumsi bahwa semua atribut tidak saling bergantungan, yang berarti nilai satu atribut tidak mempengaruhi nilai atribut lainnya. Metode *Naive Bayes* menggunakan analisis statistik untuk menentukan probabilitas awal dari setiap kelas berdasarkan data pelatihan. Probabilitas dari setiap parameter ditentukan berdasarkan probabilitas awal [10]. Secara matematis *Naïve Bayes* dapat dituliskan pada persamaan (2) berikut:

$$P(H|E) = \frac{P(E|H) \times P(H)}{P(E)} \quad (2)$$

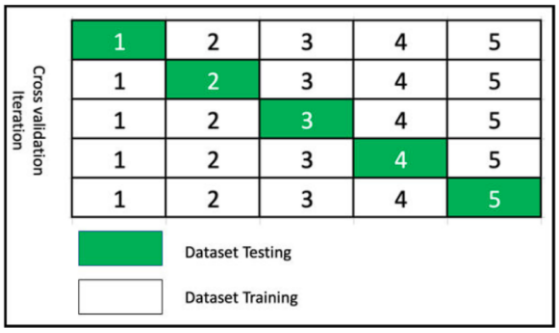
Keterangan
$P(H|E)$: Probabilitas posisional dari probabilitas (probabilitas bersyarat)
$P(E|H)$: Probabilitas parameter E pada hipotesis H
$P(H)$: Probabilitas sebelumnya (sebelumnya) hipotesis H
$P(E)$: Probabilitas awal (sebelumnya) parameter E

2.3. *Data Mining*

Knowledge Discovery in Database (KDD) adalah proses untuk mengekstrak pola dan informasi dari data yang tersimpan dalam database. Pola dan informasi ini dapat digunakan untuk membangun basis pengetahuan yang dapat digunakan untuk mendukung pengambilan keputusan [11]. Proses tahapan pada KDD seperti *selection, preprocessing, transformation, data mining, evaluation,* dan *knowledge*.

2.4. *K-Fold Cross Validation*

K-fold Cross Validation merupakan metode sampling yang umumnya diterapkan untuk membagi dataset menjadi beberapa bagian. Teknik ini sering digunakan untuk menilai atau memvalidasi tingkat akurasi suatu model yang dikembangkan berdasarkan suatu dataset. Tujuan utamanya adalah untuk mengoptimalkan waktu komputasi sambil tetap menjaga akurasi estimasi model. Prosesnya dilakukan dengan membagi data secara acak ke dalam k kelompok, di mana k merupakan jumlah lipatan atau bagian. Pembagian ini dilakukan sedemikian rupa sehingga setiap kelompok memiliki ukuran yang relatif setara, seperti pada dataset asal [12].



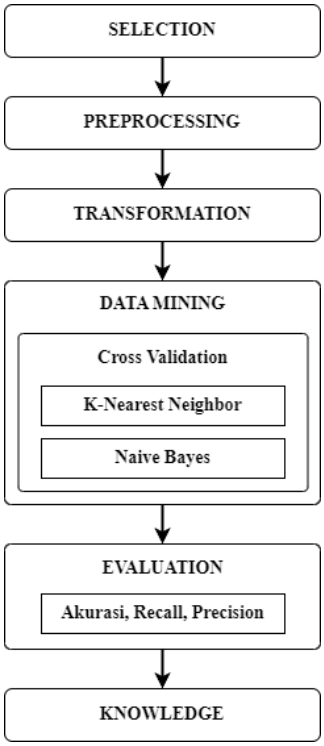
Gambar 1. K-5 *fold cross validation* [13]

Pada gambar 1, merupakan cross validation dengan jumlah lipatan sebanyak 5 *fold*.

3. METODE PENELITIAN

3.1. Metode Penelitian

Metode tahapan penelitian pada penelitian ini menggunakan *Knowledge Discovery in Database* (KDD) meliputi tahapan seperti pada gambar 2, berikut.



Gambar 2. Tahapan metode penelitian

Pada gambar 2, merupakan tahapan dari metode penelitian yang dilakukan pada penelitian ini.

3.2. Sumber Data

Penelitian ini menggunakan dataset *World Air Quality Index by City and Coordinates* yang diperoleh dari situs *Kaggle* dengan jumlah data sebanyak 16695 data. Dataset ini berisi informasi mengenai tingkat polusi udara di berbagai negara di dunia beserta koordinat lintang dan bujur kota. Dataset ini terdiri dari 14 kolom yang mencakup atribut-atribut seperti *country, city, AQI value, AQI category, AQI CO value, AQI CO category, AQI NO2 value, AQI NO2 category, AQI Ozone value, AQI Ozone category, AQI PM2.5 value,* dan *AQI PM2.5 category* dengan format csv [14].

3.3. Teknik Pengumpulan Data

Dalam penelitian ini, data akan dikumpulkan secara sekunder dari situs web Kaggle. Kata kunci yang digunakan untuk mencari data yang relevan adalah *Air Quality Index (AQI)*.

3.4. Teknik Analisis Data

Data yang telah dipilih kemudian diproses menggunakan metode *Knowledge Discovery in Database* (KDD).

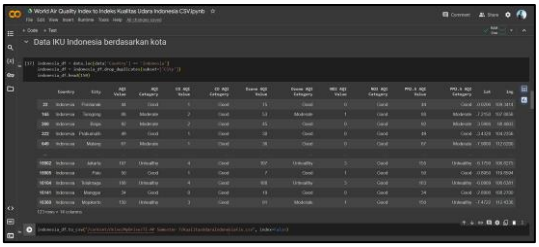
Tahapan pada *Knowledge Discovery in Database* (KDD):

1. *Selection*
   Tahapan *selection* dilakukan untuk pemilihan data yang akan digunakan.
2. *Preprocessing*
   Tahapan *preprocessing* dilakukan untuk proses melengkapi data dan menjaga konsistensi data.
3. *Transformation*
   Tahapan *transformation* dilakukan untuk mempermudah dan memperbaiki data agar sesuai dengan teknik *data mining*.
4. *Data Mining*
   Tahapan *data mining* menggunakan K-5 fold *cross validation* dengan melakukan percobaan K-2 *fold*, K-3 *fold*, K-4 *fold*, dan K-5 *fold* pada metode *K-Nearest Neighbor* dan *Naïve Bayes*.
5. *Evaluation*
   Tahapan *evaluation* dilakukan untuk evaluasi performa dari model.
6. *Knowledge*
   Tahapan terahkir adalah *knowledge* untuk memahami makna yang diperoleh dari model.

4. HASIL DAN PEMBAHASAN

4.1. *Selection*

Pada tahapan ini dilakukan proses data *selection* yang bertujuan untuk memilih data yang akan digunakan dalam proses klasifikasi tingkat kualitas udara kota di Indonesia dengan tools *RapidMiner*. Dataset yang digunakan pada penelitian ini adalah dataset *World Air Quality Index by City and Coordinates* dari situ *Kaggle* yang berjumlah 16695 data. Sementara itu, data yang digunakan pada proses *data mining* adalah data dari kota yang berasal dari Indonesia. Maka dilakukan pemisahan data dari dataset tersebut menggunakan *google colaboratory* dengan memisahkan data indeks kualitas udara berdasarkan kota dari negara Indonesia kemudian disimpan menjadi dataset baru dengan format csv seperti pada gambar 2 berikut.

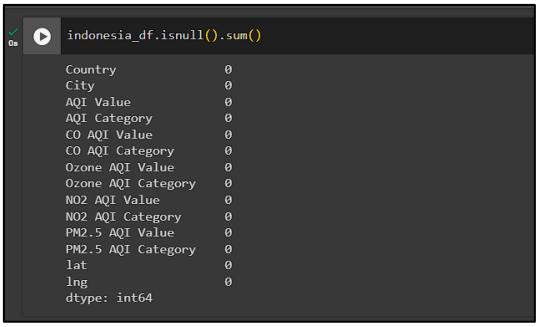


Gambar 3. Memisahkan data kualitas udara berdasarkan kota di Indonesia

Pada gambar 3, data indek kualitas udara berdasarkan kota di Indonesia berjumlah 123 data dan 14 kolom atribut.

4.2. *Preprocessing*

Pada tahap ini, data Indeks Kualitas Udara kota di Indonesia dari dataset *World Air Quality Index by City and Coordinates* yang telah dipisahkan sebelumnya dibersihkan dari *missing value*, data ganda, kesalahan data, dan data yang tidak relevan dengan menggunakan *google colaboratory* seperti pada gambar 3 berikut.



Gambar 4. Informasi *missing value* dari dataset

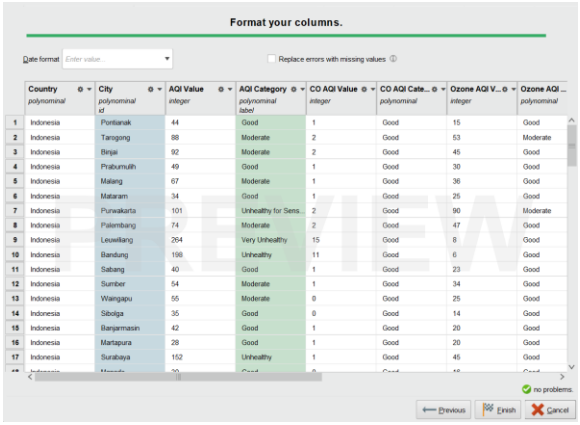
Pada gambar 4, berisi informasi dari jumlah missing value pada dataset untuk setiap kolom.

4.3. *Transformation*

Atribut dari data sudah sesuai untuk dilakukan proses klasifikasi menggunakan *K-Nearest Neighbor* dan *Naïve Bayes* pada tahap selanjutnya dengan *RapidMiner*, sehingga tahapan ini tidak dilakukan proses *transformation* pada dataset.

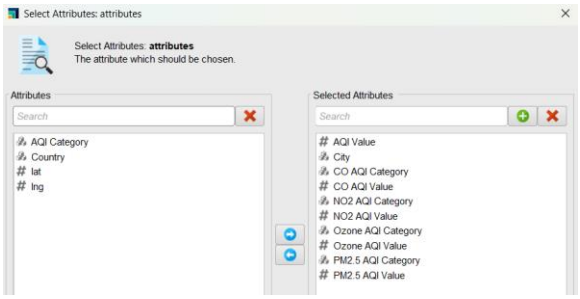
4.4. *Data Mining*

Pada tahapan ini dilakukan proses data mining dengan metode *K-Nearest Neighbor* dan *Naïve Bayes* untuk klasifikasi tingkat kualitas udara kota di Indonesia. Langkah pertama yang dilakukan adalah mengimport data yang digunakan dengan operator *Read CSV* sekaligus dilakukan pemberian label pada attribut seperti pada gambar 4 berikut.



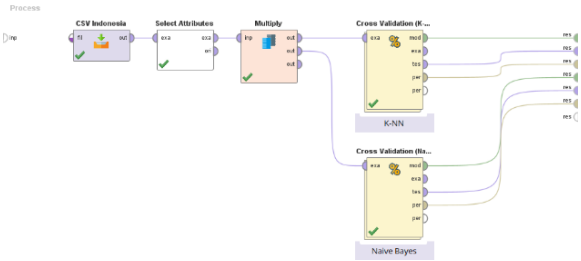
Gambar 5. Import dataset serta pemberian label dan id pada *RapidMiner*

Berdasarkan gambar 5, atribut *AQI Category* dijadikan sebagai label dan atribut *City* dijadikan sebagai *id*. Untuk memilih atribut dari data yang akan digunakan pada model, hubungkan operator *Read CSV* dengan operator *Select Attributes*. Atribut yang digunakan yaitu *AQI value, city, CO AQI value, CO AQI category, NO2 AQI value, NO2 AQI category, Ozone AQI value, Ozone AQI category, PM2.5 AQI value* dan *PM2.5 AQI category* yang dapat pada gambar 6 berikut.



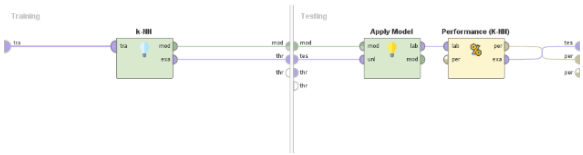
Gambar 6. Atribut yang digunakan

Langkah selanjutnya adalah menghubungkan operator *Select Attributes* ke operator *Multiply* untuk menduplikasi dataset menjadi dua. Kemudian hubungkan operator *Multiply* dengan operator *Cross-validation* untuk metode *K-Nearest Neighbor* dan *Naïve Bayes* seperti pada gambar 7 berikut.



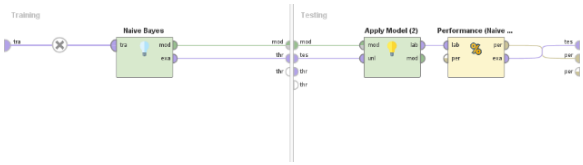
Gambar 7. Model klasifikasi tingkat kualitas udara di Indonesia dengan cross-validation

Pada gambar 7, operator *Cross-validation* metode *K-Nearest Neighbor* dan *Naïve Bayes* terbagi menjadi *training* dan *testing* di dalam prosesnya dapat dilihat pada gambar 8 berikut.



Gambar 8. *Cross-validation* metode *K-Nearest Neighbor*

Berdasarkan gambar 8, merupakan proses di dalam operator *Cross-validation* pada metode *K-Nearest Neighbor* yang terdiri dari bilah *training* di sebelah kiri dan *testing* di sebelah kanan. Pada bilah training terdapat operator k-NN, kemudian pada bilah testing terdapat operator *Apply Model* dan *Performance (Classification)*. Pada operator k-NN nilai K yang digunakan pada penelitian ini adalah k-1 dengan *measure types MixedMeasures* dan mixed measure *MixedEuclideanDistance.*



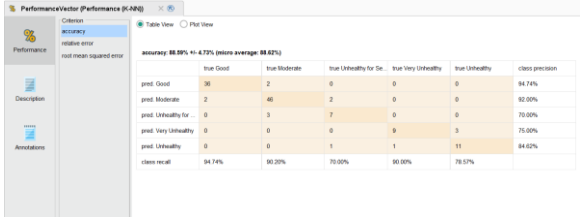
Gambar 9. *Cross-validation* metode *Naïve Bayes*

Berdasarkan gambar 9, merupakan proses di dalam operator *Cross-validation* pada metode *Naïve Bayes* yang terdiri dari bilah *training* di sebelah kiri dan *testing* di sebelah kanan. Pada bilah training terdapat operator *Naïve Bayes*, kemudian pada bilah testing terdapat operator *Apply Model* dan *Performance (Classification)*.

Pada penelitian ini nilai *number of folds* pada operator *Cross-validation* yang digunakan adalah *K*-5 *fold cross validation* dengan melakukan percobaan K-2 *fold*, K-3 *fold*, K-4 *fold*, dan K-5 *fold* untuk mencari nilai *number of folds* terbaik dari metode *K-Nearest Neighbor* dan *Naïve Bayes*.

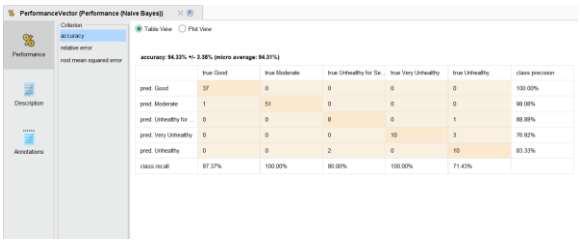
Hasil dari akurasi untuk setiap nilai *number of folds* adalah sebagai berikut:

1. Hasil Akurasi K-2 *fold*
   Akurasi *Performance (Classification)* metode *K-Nearest Neighbor* dan *Naïve Bayes* dapat dengan menggunakan K-2 fold dapat dilihat pada gambar 10 dan 11 berikut.



Gambar 10. Akurasi K-2 *fold* metode *K-Nearest Neighbor*

Berdasarkan gambar 10, metode *K-Nearest Neighbor* dengan menggunakan K-2 *fold* memperoleh nilai akurasi 88.59%.



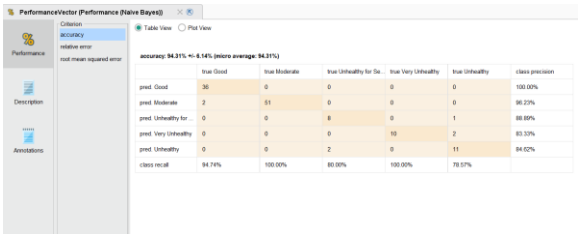
Gambar 11. Akurasi K-2 *fold* metode *Naïve Bayes*

Berdasarkan gambar 11, metode *Naïve Bayes* dengan menggunakan K-2 *fold* memperoleh nilai akurasi 94.33%.

2. Hasil Akurasi K-3 *fold*
   Akurasi *Performance (Classification)* metode *K-Nearest Neighbor* dan *Naïve Bayes* dapat dengan menggunakan K-3 fold dapat dilihat pada gambar 12 dan 13 berikut.

Gambar 12. Akurasi K-3 *fold* metode *K-Nearest Neighbor*

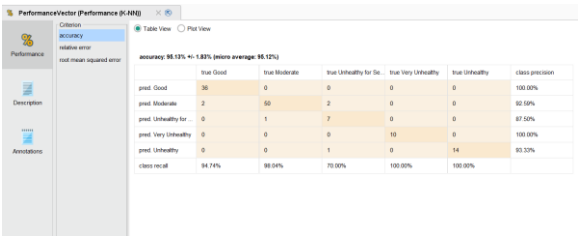
Berdasarkan gambar 12, metode K-Nearest Neighbor dengan menggunakan K-3 fold memperoleh nilai akurasi 93.50%.



Gambar 13. Akurasi K-3 *fold* metode *Naïve Bayes*

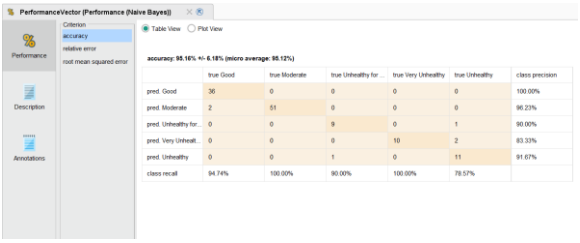
Berdasarkan gambar 13, metode *Naïve Bayes* dengan menggunakan K-3 *fold* memperoleh nilai akurasi 94.31%.

3. Hasil Akurasi K-4 *fold*
   Akurasi *Performance (Classification)* metode *K-Nearest Neighbor* dan *Naïve Bayes* dapat dengan menggunakan K-4 fold dapat dilihat pada gambar 14 dan 15 berikut.



Gambar 14. Akurasi K-4 *fold* metode *K-Nearest Neighbor*

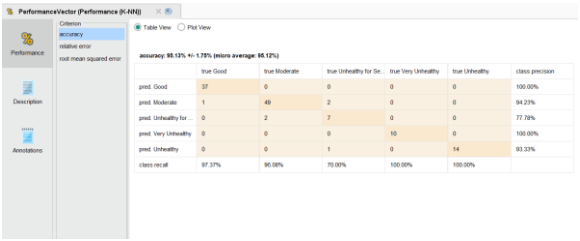
Berdasarkan gambar 14, metode *K-Nearest Neighbor* dengan menggunakan K-4 *fold* memperoleh nilai akurasi 95.13%.



Gambar 15. Akurasi K-4 *fold* metode *Naïve Bayes*

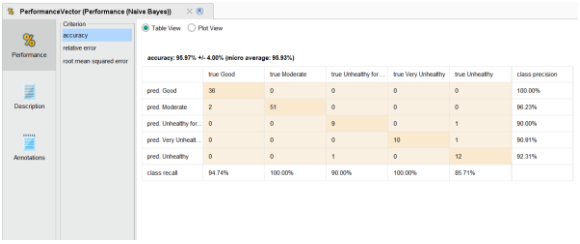
Berdasarkan gambar 15, metode *Naïve Bayes* dengan menggunakan K-4 *fold* memperoleh nilai akurasi 95.16%.

4. Hasil Akurasi K-5 *fold*
   Akurasi *Performance (Classification)* metode *K-Nearest Neighbor* dan *Naïve Bayes* dapat dengan menggunakan K-5 fold dapat dilihat pada gambar 16 dan 17 berikut.



Gambar 16. Akurasi K-5 *fold* metode *K-Nearest Neighbor*

Berdasarkan gambar 16, metode *K-Nearest Neighbor* dengan menggunakan K-5 *fold* memperoleh nilai akurasi 95.13%.



Gambar 17. Akurasi K-5 *fold* metode *Naïve Bayes*

Berdasarkan gambar 17, metode *Naïve Bayes* dengan menggunakan K-5 *fold* memperoleh nilai akurasi 95.97%.

4.5. *Evaluation*

Hasil evaluasi nilai akurasi dengan K-5 *fold cross validation* yang diperoleh dari metode *K-Nearest Neighbor* dan *Naïve Bayes* dapat dilihat pada tabel 1 berikut.

Tabel 1. Akurasi K-5 *fold cross validation*

Number of fold	**Akurasi K-Nearest Neighbor (%)**	**Akurasi Naïve Bayes (%)**
K-2 *fold*	88.59	94.33
K-3 *fold*	93.50	94.31
K-4 *fold*	95.13	95.16
K-5 *fold*	95.13	95.97

Pada tabel 1, merupakan hasil nilai akurasi dari setiap percobaan menggunakan K-5 *fold cross validation* mulai dari K-2 *fold* sampai K-5 *fold* untuk masing-masing metode.

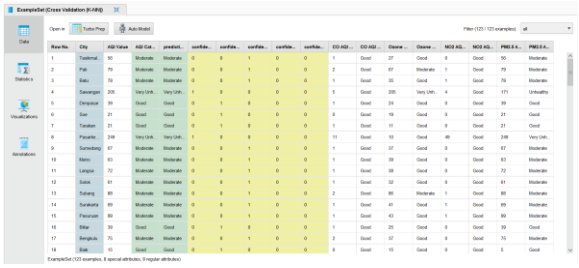
Tabel 2. Nilai *recall* dan *precision* K-5 *fold K-Nearest Neighbor* dan *Naïve Bayes*

Kelas	***K-Nearest Neighbor (%)***		***Naïve Bayes (%)***	
	Recall	*Precision*	*Recall*	*Precision*
Good	97.37	100	94.74	100
Moderate	96.08	94.23	100	96.23
Unhealthy for Sensitive Groups	70.00	77.78	90.00	90.00
Unhealthy	100	93.33	85.71	92.31
Very Unhealthy	100	100	100	90.91

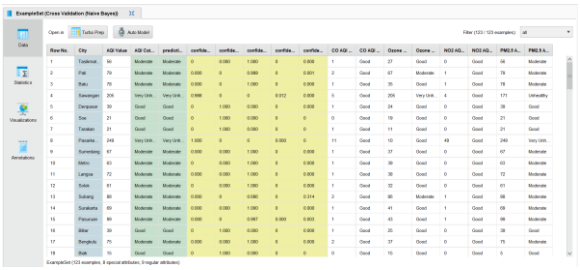
Pada tabel 2, merupakan nilai dari *recall* dan *precision* untuk setiap kelas *AQI Category* dari metode *K-Nearest Neighbor* dan *Naïve Bayes* pada K-5 *fold*.

4.6. *Knowledge*

Berdasarkan tahapan yang telah dilakukan pada penelitian ini menghasilkan tabel hasil klasifikasi tingkat kualitas udara untuk setiap kota sebanyak 123 data dari metode *K-Nearest Neighbor* dan *Naïve Bayes* yang dapat dilihat pada gambar 17 dan 18 berikut.

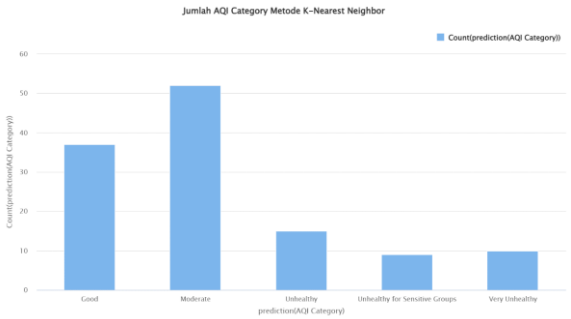


Gambar 18. Hasil klasifikasi metode *K-Nearest Neighbor* pada *RapidMiner*



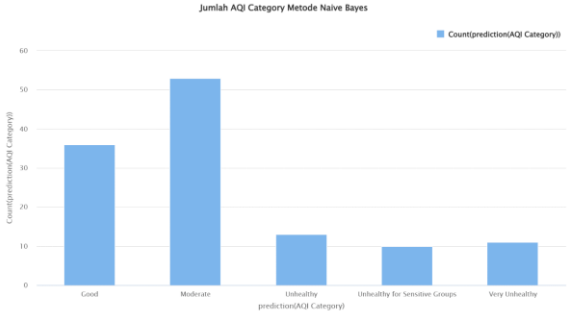
Gambar 19. Hasil klasifikasi metode *Naïve Bayes* pada *RapidMiner*

Berdasarkan gambar 18 dan 19, kolom yang berwarna hijau merupakan label dan prediksi kategori kelas dari *AQI Category* yang dihasilkan dari metode *K-Nearest Neighbor* dan *Naïve Bayes*, dan untuk kolom yang berwarna kuning merupakan nilai *confidence* untuk setiap kelas-nya.



Gambar 20. Jumlah *AQI Category* metode *K-Nearest Neighbor* pada *RapidMiner*

Pada gambar 20, merupakan jumlah dari hasil prediction *AQI Category* untuk tiap kelas pada metode *K-Nearest Neighbor*.



Gambar 21. Jumlah *AQI Category* metode *Naïve Bayes* pada *RapidMiner*

Pada gambar 21, merupakan jumlah dari hasil prediction *AQI Category* untuk tiap kelas pada metode *Naïve Bayes*.

Penelitian ini dilakukan penerapan klasifikasi data indeks kualitas udara kota di Indonesia dengan menggunakan metode *K-Nearest Neighbor* dan *Naïve Bayes*, KDD (*Knowledge Discovery in Database*) digunakan sebagai metode analisis data, *RapidMiner* digunakan sebagai *tools* untuk membuat model dan penerapan klasifikasinya. Beberapa tahapan dilakukan untuk analisis data yang terdiri dari tahapan *Selection*,

Preprocessing, Transformation, Data mining, Evaluation, dan *Knowledge.* Pada tahap *data mining* menggunakan K-5 *fold cross validation* untuk masing-masing metode *K-Nearest Neighbor* dan *Naïve Bayes* dari data yang telah diseleksi pada tahap *Selection*. Hasil yang diperoleh dari model dengan melakukan percobaan K-2 *fold*, K-3 *fold*, K-4 *fold*, dan K-5 *fold* menghasilkan nilai akurasi, *recall, precision* terbaik pada K-5 *fold* dan tabel klasifikasi kualitas udara kota di Indonesia untuk masing-masing metode *K-Nearest Neighbor* dan *Naïve Bayes*. Metode *K-Nearest Neighbor* menghasilkan akurasi 95.13% dan metode *Naïve Bayes* menghasilkan akurasi 95.97% pada K-5 *fold*. Kemudian untuk nilai *recall* dan *precision* untuk K-5 *fold* berdasarkan tabel 2 di atas, metode *Naïve Bayes* menghasilkan nilai *recall* dan *precision* di atas 85% untuk semua kelas lebih dari metode *K-Nearest Neighbor* yang memiliki salah satu nilai *recall* dan *precision* di bawah 80% pada kelas *Unhealthy for Sensitive Groups.*

5. KESIMPULAN DAN SARAN

Berdasarkan hasil dari proses klasifikasi data Indek Kualitas Udara kota di Indonesia menggunakan metode *K-Nearest Neighbor* dan *Naïve Bayes* dengan K-5 *fold cross validation* untuk percobaan K-2 *fold*, K-3 *fold*, K-4 *fold* dan K-5 *fold* maka dapat disimpulkan bahwa nilai akurasi yang terbaik adalah K-5 *fold*, pada metode *K-Nearest Neighbor* menghasilkan nilai akurasi 95.13%, sementara untuk metode *Naïve Bayes* menghasilkan nilai akurasi 95.97%. Sehingga metode *K-Nearest Neighbor* dan *Naïve Bayes* menghasilkan performa yang sangat baik karena memiliki nilai akurasi diatas 95%. Kemudian hasil nilai *recall* dan *precision* untuk K-5 *fold* menghasilkan nilai multi kelas untuk metode *K-Nearest Neighbor* dan *Naïve Bayes* meliputi kelas *Good, Moderate, Unhealthy for Sensitive Groups, Unhealthy*, dan *Very Unhealthy*. Metode *Naïve Bayes* menghasilkan nilai *recall* dan *precision* di atas 85% untuk semua kelas lebih dari metode *K-Nearest Neighbor* yang memiliki salah satu nilai *recall* dan *precision* di bawah 80% pada kelas *Unhealthy for Sensitive Groups.* Berdasarkan hasil yang diperoleh dari nilai akurasi, *recall*, dan *precision* maka metode *Naïve Bayes* memiliki hasil yang lebih baik dibandingkan metode *K-Nearest Neighbor* untuk mengklasifikasi data Indeks Kualitas Udara kota di Indonesia.

Dalam rangka meningkatkan dan memperbaiki kekurangan untuk penyempurnaan dari penelitian ini maka diberikan beberapa saran sebagai berikut: Penelitian ini dapat dikembangkan lagi dengan menggunakan algoritma atau metode klasifikasi lain seperti *Support Vector Machine, C4.5, Decision Tree, Neural Networks*, dan lain-lain. Pada penelitian selanjutnya, lebih baik untuk menggunakan data yang lebih banyak.

DAFTAR PUSTAKA

[1] S. Nurjanah, A. M. Siregar, and D. S. Kusumaningrum, "Penerapan Algoritma K–Nearest Neighbor (KNN) Untuk Klasifikasi Pencemaran Udara Di Kota Jakarta," *Scientific Student Journal for Information, Technology and Science*, vol. 1, No. 2, pp. 71–76, 2020, [Online]. Available: http://journal.ubpkarawang.ac.id/mahasiswa/index.php/ssj/article/download/14/12

[2] A. Amalia, A. Zaidiah, and I. N. Isnainiyah, "Prediksi Kualitas Udara Menggunakan Algoritma K-Nearest Neighbor," *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 7, No. 2, pp. 496–507, 2022, [Online]. Available: https://www.jurnal.stkippgritulungagung.ac.id/index.php/jipi/article/view/2843

[3] IQAir, "Kualitas udara di Indonesia." Accessed: Aug. 04, 2023. [Online]. Available: https://www.iqair.com/id/indonesia

[4] A. A. H. Kirono, I. Asror, and Y. F. A. Wibowo, "Klasifikasi Tingkat Kualitas Udara Dki Jakara Dengan Algoritma Naive Bayes," *eProceedings of Engineering*, vol. 9, No. 3, pp. 1962–1969, 2022, [Online]. Available: https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/18002

[5] A. Nugroho, I. Asror, and Y. F. A. Wibowo, "Klasifikasi Tingkat Kualitas Udara DKI Jakarta Berdasarkan Open Government Data Menggunakan Algoritma Random Forest," *eProceedings of Engineering*, vol. 10, No. 2, pp. 1824–1834, 2023, [Online]. Available: https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/20030

[6] D. D. Purwanto and E. S. Honggara, "Klasifikasi Kategori Hasil Perhitungan Indeks Standar Pencemaran Udara dengan Gausian Naïve Bayes (Studi Kasus: ISPU DKI Jakarta 2020)," *INSYST: Journal of Intelligent System and Computation*, vol. 4, No. 2, pp. 102–108, 2022, doi: 10.52985/insyst.v4i2.259.

[7] A. D. Wiranata, S. Soleman, Irwansyah, I. K. Sudaryana, and 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, No. 9, pp. 95–100, 2023, doi: 10.37365/jti.v9i1.164.

[8] D. N. K. Hardani, T. Luthfianto, and M. T. Tamam, "Identify The Authenticity of Rupiah Currency Using K Nearest Neighbor (K-NN) Algorithm," *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika*, vol. 5, no. 1, Jul. 2019, doi: 10.26555/jiteki.v5i1.13324.

[9] M. Gunawan, M. Zarlis, and R. Roslina, "Analisis Komparasi Algoritma Naïve Bayes dan K-Nearest Neighbor Untuk Memprediksi

Kelulusan Mahasiswa Tepat Waktu," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 5, No. 2, pp. 513–523, 2021, doi: 10.30865/mib.v5i2.2925.

[10] A. Y. P. Yusuf and R. Sari, "Implementasi Algoritma Naïve Bayes Untuk Klasifikasi Pemahaman Program MBKM Bagi Mahasiswa," *Journal of Informatic and Information Security (JIFORTY)*, vol. 3, No. 2, pp. 171–180, 2022, [Online]. Available: https://ejurnal.ubharajaya.ac.id/index.php/jiforty/article/view/1713

[11] G. Gustientiedina, M. H. Adiya, and Y. Desnelita, "Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan Pada RSUD Pekanbaru," *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 5, no. 1, pp. 17–24, Apr. 2019, doi: 10.25077/TEKNOSI.v5i1.2019.17-24.

[12] N. Fibriyanti Arminda, N. Sulistiyowati, and T. Nur Padilah, "IMPLEMENTASI ALGORITMA MULTINOMIAL NAIVE BAYES PADA ANALISIS SENTIMEN TERHADAP ULASAN PENGGUNA APLIKASI BRIMO," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 7, no. 3, pp. 1817–1822, Nov. 2023, doi: 10.36040/jati.v7i3.7012.

[13] D. Hindarto and H. Santoso, "Performance Comparison of Supervised Learning Using Non-Neural Network and Neural Network," *Jurnal Nasional Pendidikan Teknik Informatika (JANAPATI)*, vol. 11, no. 1, p. 49, Apr. 2022, doi: 10.23887/janapati.v11i1.40768.

[14] Aditya Ramachandran, "World Air Quality Index by City and Coordinates." Accessed: Dec. 29, 2023. [Online]. Available: https://www.kaggle.com/datasets/adityaramachandran27/world-air-quality-index-by-city-and-coordinates