

BAB V

HASIL DAN PEMBAHASAN

Bab ini akan memaparkan hasil dari penelitian yang telah dilakukan serta pembahasan dari hasil tersebut. Berikut merupakan uraian hasil penelitian ini.

5.1. Hasil dan Pembahasan Perencanaan

Pada bab 4 telah dijelaskan perencanaan dari penelitian ini yang digambarkan dalam bentuk diagram alir. Dengan mengacu pada perencanaan tersebut, berikut merupakan hasil penelitian yang akan dijelaskan sebagai berikut.

5.1.1. Data *Pre-processing*

Pada bab ini, data memasuki tahap pertama yaitu tahap *pre-processing* sebelum dimasukkan ke dalam model. Di tahap ini data juga akan dinormalisasikan. Data yang belum di normalisasikan dapat dilihat pada Tabel 5.1.

Pada data asli yang tertera di Tabel 5.1, terdapat parameter 15 parameter. Berikut penjelasan dari masing-masing parameter.

1. Kekeruhan, merupakan tingkat kekeruhan dari air tersebut. Batas Kekeruhan yang diperbolehkan adalah maksimal 5 NTU.
2. Air Raksa, merupakan kandungan logam cair yang berwarna keperakan. Air Raksa pada data ini memiliki satuan mg/liter. Batas air raksa yang diperbolehkan hanya 0.001 mg/liter.
3. Besi (Fe), merupakan parameter kimiawi, batas kadar Besi (Fe) yang diperbolehkan maksimal 0.3 mg/l.
4. Fluorida, merupakan senyawa kimia, batas Fluorida yang diperbolehkan adalah maksimal 1.5 mg/liter.
5. Kesadahan, merupakan parameter kimiawi yang dimiliki oleh air. Batas Kesadahan yang diperbolehkan adalah maksimal 500 mg/l.
6. Krom heksavalen, merupakan senyawa yang didalamnya terkandung kromium.

7. Mangan (Mn), merupakan sebuah unsur ikatan logam perak. Batas Mangan yang diperbolehkan adalah maksimal 0.1 mg/l.
8. Nitrat, merupakan sebuah ion poliatomik. Batas Nitrat yang diperbolehkan adalah maksimal 50 mg/liter.
9. Nitrit, merupakan salah satu unsur yang menjadi penyusun sel organisme yakni penyusun dalam pembentukan protoplasma. Batas Nitrit yang diperbolehkan adalah maksimal 3 mg/liter.
10. Seng (Zn), merupakan timah sari, atau sebuah logam putih kebiruan berkilau. Batas Seng yang diperbolehkan adalah maksimal 3 mg/l.
11. Sulfat, merupakan sebuah garam atau ester asam sulfat. Batas Sulfat yang diperbolehkan adalah maksimal 250 mg/l.
12. Detergen, merupakan bahan yang digunakan dalam membantu proses pembersihan.
13. KMnO_4 , merupakan kepanjangan dari Kalium permanganat. KMnO_4 adalah sebuah senyawa anorganik yang mengandung mangan, oksigen dan kalium.
14. Total Coliform, merupakan bakteri yang ada pada saluran pencernaan hewan dan manusia. Bakteri ini dapat ditemukan pada kotoran hewan, manusia, tanaman, dan tanah. Batas Total Coliform yang diperbolehkan adalah maksimal 0 Jumlah dari per 100 ml sampel.
15. E. Coli, merupakan salah satu jenis spesies dari bakteri negatif. E. Coli dapat ditemukan di dalam usus manusia dan hewan sehat. Batas E. Coli yang diperbolehkan adalah 0 Jumlah dari per 100 ml sampel.

Tabel 5.1 Data asli

No	Kekeruhan	Air Raksa	Besi (Fe)	Fluorida	Kesadahan	Krom Heksavalen	Mangan (Mn)	Nitrat	Nitrit	Seng (Zn)	Sulfat	Detergen	KMnO4	Total Coliform	E.Coli	Kategori
0	0.46	0.0003	0.08	0.48	23.68	0	0.05	0	0.01	0.01	70.5	0.18	7.33	900	0	2
1	0.39	0.0003	0.14	0.51	107.46	0	0.1	0.88	0.03	0.01	40.7	0.06	13.85	100	0	1
2	4.13	0.0003	0.16	0.61	78.01	0	0.05	0	0.01	0.01	296.61	0.04	14.69	600	0	1
3	7.89	0.0003	0.35	0.35	92.14	0	0.05	0.52	0.01	0.01	39.55	0.06	4.53	100	0	1
4	1.91	0.0003	0.08	0.78	151.64	0	0.1	0.26	0.01	0.01	62.81	0.07	18.15	30	400	3
5	1.85	0.0003	0.08	0.23	83.93	0	0.1	0.37	0	0.01	73.56	0.15	2.47	1.7	100	2
6	0.71	0.0003	0.08	0.25	109.52	0	0.05	1.6	0	0.01	25.35	0.07	2.93	200	0	1
7	0.67	0.0003	0.08	0.08	51.39	0	0.08	4.05	0.03	0.01	31.65	0.11	1.43	0	0	1
8	0.46	0.0003	0.08	0.26	21.69	0	0.05	0.2	0	0.01	5.3	0.15	4.56	50	0	1
9	0.28	0.0003	0.08	0.1	68.66	0	0.05	2.89	0.01	0.01	4.64	0.11	2.51	200	0	1

Pada data asli terdapat 16 fitur yang terdiri dari informasi mengenai kekeruhan, air raksa, besi (Fe), fluorida, kesadahan, krom heksavalen, mangan (Mn), nitrat, nitrit, seng (Zn), sulfat, detergen, KMnO₄, total coliform, E.Coli, dan kategori. Data tersebut merupakan data dari pengambilan sampel air di provinsi DKI Jakarta pada tahun 2018 – 2019. Pada Tabel 4.2 di atas, data pada fitur kategori telah diubah dari yang sebelumnya data non-numerik menjadi data numerik. Perubahan data ini telah dijelaskan pada Tabel 4.1.

Pada penelitian ini, normalisasi dilakukan menggunakan Teknik *min-max normalization*. Normalisasi dilakukan dengan cara mengurangi setiap fitur dengan nilai minimal dari fitur tersebut. Sehingga didapatkan hasil dengan rentang nilai 0 – 1 (Saputra & Kristiyanti, 2022). Teknik normalisasi ini memiliki rumus yang dapat dilihat pada (4).

$$X_{new} = \frac{X_{old} - X_{min}}{X_{max} - X_{min}} \quad (4)$$

Rumus (4) di atas memiliki keterangan sebagai berikut.

- a) X = merupakan fitur yang dihitung.
- b) X_{new} = merupakan nilai fitur baru yang dihasilkan.
- c) X_{old} = merupakan nilai fitur lama yang dihasilkan.
- d) X_{min} = merupakan nilai minimal yang ada pada fitur tersebut.
- e) X_{max} = merupakan nilai maksimal yang ada pada fitur tersebut.

Hasil dari proses normalisasi menggunakan *min-max normalization* dapat dilihat pada Tabel 5.2.

Tabel 5.2 Normalisasi data

No	Kekeruhan	Air Raksa	Besi (Fe)	Fluorida	Kesadahan	Krom Heksa valen	Mangan (Mn)	Nitrat	Nitrit	Seng (Zn)	Sulfat	Detergen	KMnO4	Total Coliform	E.Coli
0	0.00494	0.000429	0.005333	0.053333	0.003242	0	0.00344	0	5E-06	0.00285	0.129868	0.2	0.029438	6.87E-08	0.00E+00
1	0.00419	0.000429	0.009333	0.056667	0.015752	0	0.00687	0.16327	2.3E-05	0.00285	0.074973	0.06667	0.055622	7.63E-09	0.00E+00
2	0.04438	0.000429	0.010667	0.067778	0.011354	0	0.00344	0	5E-06	0.00285	0.546384	0.04444	0.058996	4.58E-08	0.00E+00
3	0.08478	0.000429	0.023333	0.038889	0.013464	0	0.00344	0.09648	5E-06	0.00285	0.072855	0.06667	0.018193	7.63E-09	0.00E+00
4	0.02052	0.000429	0.005333	0.086667	0.022349	0	0.00687	0.04824	1.1E-05	0.00285	0.115702	0.07778	0.072892	2.29E-09	5.71E-08
5	0.01988	0.000429	0.005333	0.025556	0.012238	0	0.00687	0.06865	2E-06	0.00285	0.135505	0.16667	0.00992	1.30E-10	1.43E-08
6	0.00763	0.000429	0.005333	0.027778	0.016059	0	0.00344	0.29685	2E-06	0.00285	0.046697	0.07778	0.011767	1.53E-08	0.00E+00
7	0.0072	0.000429	0.005333	0.008889	0.007379	0	0.0055	0.75139	2.6E-05	0.00285	0.058302	0.12222	0.005743	0.00E+00	0.00E+00
8	0.00494	0.000429	0.005333	0.028889	0.002945	0	0.00344	0.03711	2E-06	0.00285	0.009763	0.16667	0.018313	3.82E-09	0.00E+00
9	0.00301	0.000429	0.005333	0.011111	0.009958	0	0.00344	0.53618	4E-06	0.00285	0.008547	0.12222	0.01008	1.53E-08	0.00E+00

5.1.2. Pembagian Data

Terdapat 1053 data yang telah melewati tahap *pre-processing*. Data tersebut selanjutnya dibagi dua menjadi data latih dan data uji. Pada pembagian data ini, dilakukan pengujian menggunakan 6 macam rasio. Hal tersebut dilakukan untuk mencari pembagian rasio terbaik dari hasil perbandingan 6 macam rasio yang diuji.

Pembagian data latih dan data uji yang dilakukan menggunakan beberapa rasio. Rasio pertama yang diuji oleh peneliti adalah 50% data dijadikan data latih dan 50% data lainnya menjadi data uji. Melalui pengujian tersebut, didapatkan hasil nilai akurasi sebesar 0.81594, dilanjutkan dengan nilai presisi sebesar 0.8467, lalu nilai recall sebesar 0.78669, dan terakhir nilai F1-score sebesar 0.80983. Hasil ditunjukkan pada Tabel 5.3.

Tabel 5.3 *Random forest* rasio pembagian 50:50

Rasio	n_estimators	Presisi	Recall	Akurasi	F1-Score
50:50	100	0.8467	0.78669	0.81594	0.80983

Selanjutnya Pembagian data latih dan data uji yang dilakukan menggunakan rasio lainnya. Rasio kedua yang diuji oleh peneliti adalah 60% data dijadikan data latih dan 40% data lainnya menjadi data uji. Melalui pengujian tersebut, didapatkan hasil nilai akurasi sebesar 0.8436, dilanjutkan dengan nilai presisi sebesar 0.87844, lalu nilai recall sebesar 0.79388, dan terakhir nilai F1-score sebesar 0.82835. Hasil ditunjukkan pada Tabel 5.4.

Tabel 5.4 *Random forest* rasio pembagian 60:40

Rasio	n_estimators	Presisi	Recall	Akurasi	F1-Score
60:40	100	0.87844	0.79388	0.8436	0.82835

Pembagian data latih dan data uji yang dilakukan menggunakan rasio lainnya. Rasio ketiga yang diuji oleh peneliti adalah 70% data dijadikan data latih dan 30% data lainnya menjadi data uji. Melalui pengujian tersebut, didapatkan hasil nilai akurasi sebesar 0.8481, dilanjutkan dengan nilai presisi sebesar 0.8576, lalu nilai recall sebesar 0.79152, dan terakhir nilai F1-score sebesar 0.82025. Hasil ditunjukkan pada Tabel 5.5.

Tabel 5.5 *Random forest* rasio pembagian 70:30

Rasio	n_estimators	Presisi	Recall	Akurasi	F1-Score
70:30	100	0.8576	0.79152	0.8481	0.82025

Rasio keempat yang diuji oleh peneliti adalah 75% data dijadikan data latih dan 25% data lainnya menjadi data uji. Melalui pengujian tersebut, didapatkan hasil nilai akurasi sebesar 0.85227, dilanjutkan dengan nilai presisi sebesar 0.88112, lalu nilai recall sebesar 0.79891, dan terakhir nilai F1-score sebesar 0.83351. Hasil ditunjukkan pada Tabel 5.6.

Tabel 5.6 *Random forest* rasio pembagian 75:25

Rasio	n_estimators	Presisi	Recall	Akurasi	F1-Score
75:25	100	0.88112	0.79891	0.85227	0.83351

Rasio kelima yang diuji oleh peneliti adalah 80% data dijadikan data latih dan 20% data lainnya menjadi data uji. Melalui pengujian tersebut, didapatkan hasil nilai akurasi sebesar 0.83886, dilanjutkan dengan nilai presisi sebesar 0.82466, lalu nilai recall sebesar 0.77618, dan terakhir nilai F1-score sebesar 0.79656. Hasil ditunjukkan pada Tabel 5.7.

Tabel 5.7 *Random forest* rasio pembagian 80:20

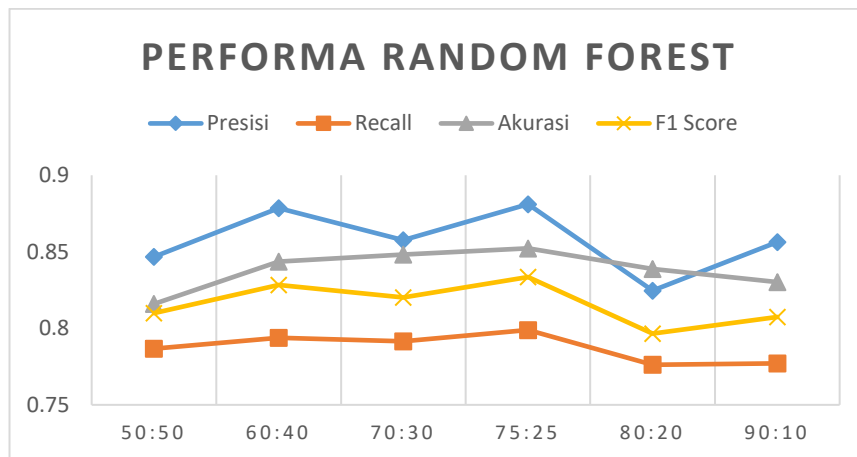
Rasio	n_estimators	Presisi	Recall	Akurasi	F1-Score
80:20	100	0.82466	0.77618	0.83886	0.79656

Rasio kelima yang diuji oleh peneliti adalah 90% data dijadikan data latih dan 10% data lainnya menjadi data uji. Melalui pengujian tersebut, didapatkan hasil nilai akurasi sebesar 0.83019, dilanjutkan dengan nilai presisi sebesar 0.85628, lalu nilai recall sebesar 0.77718, dan terakhir nilai F1-score sebesar 0.80737. Hasil ditunjukkan pada Tabel 5.8.

Tabel 5.8 *Random forest* rasio pembagian 90:10

Rasio	n_estimators	Presisi	Recall	Akurasi	F1-Score
90:10	100	0.85628	0.77718	0.83019	0.80737

Hasil pembagian data untuk pemodelan algoritma *random forest* ditunjukkan pada Gambar 5.1. Melalui grafik di tersebut dapat dilihat bahwa rasio keempat yaitu 75% data latih dan 25% data uji menghasilkan nilai akurasi tertinggi performa lainnya. Maka dari itu peneliti memutuskan untuk menggunakan rasio 75% data latih dan 25% data uji untuk *random forest*.



Gambar 5.1 Perbandingan pembagian data *random forest*

Selanjutnya untuk pembagian data latih dan data uji yang dilakukan dalam pemodelan menggunakan *naive bayes* juga menggunakan 6 perbandingan rasio. Rasio pertama yang diuji oleh peneliti adalah 50% data dijadikan data latih dan 50% data lainnya menjadi data uji. Melalui pengujian tersebut, didapatkan hasil nilai akurasi sebesar 0.70398, dilanjutkan dengan nilai presisi sebesar 0.75423, lalu nilai recall sebesar 0.75856, dan terakhir nilai F1-score sebesar 0.73066. Hasil ditunjukkan pada Tabel 5.9.

Tabel 5.9 *Naive bayes* rasio pembagian 50:50

Rasio	Classifier	Precisi	Recall	Akurasi	F1-Score
50:50	Gaussian	0.75423	0.75856	0.70398	0.73066

Rasio kedua yang diuji oleh peneliti adalah 60% data dijadikan data latih dan 40% data lainnya menjadi data uji. Melalui pengujian tersebut, didapatkan hasil nilai akurasi sebesar 0.71564, dilanjutkan dengan nilai presisi sebesar 0.75382, lalu nilai recall sebesar 0.75807, dan terakhir nilai F1-score sebesar 0.73482. Hasil ditunjukkan pada Tabel 5.10.

Tabel 5.10 *Naive bayes* rasio pembagian 60:40

Rasio	Classifier	Presisi	Recall	Akurasi	F1-Score
60:40	Gaussian	0.75382	0.75807	0.71564	0.73482

Rasio ketiga yang diuji oleh peneliti adalah 70% data dijadikan data latih dan 30% data lainnya menjadi data uji. Melalui pengujian tersebut, didapatkan hasil nilai akurasi sebesar 0.71835, dilanjutkan dengan nilai presisi sebesar 0.74859, lalu nilai recall sebesar 0.77555, dan terakhir nilai F1-score sebesar 0.74005. Hasil ditunjukkan pada Tabel 5.11.

Tabel 5.11 *Naive bayes* rasio pembagian 70:30

Rasio	Classifier	Presisi	Recall	Akurasi	F1-Score
70:30	Gaussian	0.74859	0.77555	0.71835	0.74005

Rasio keempat yang diuji oleh peneliti adalah 75% data dijadikan data latih dan 25% data lainnya menjadi data uji. Melalui pengujian tersebut, didapatkan hasil nilai akurasi sebesar 0.7197, dilanjutkan dengan nilai presisi sebesar 0.76218, lalu nilai recall sebesar 0.7625, dan terakhir nilai F1-score sebesar 0.7434. Hasil ditunjukkan pada Tabel 5.12.

Tabel 5.12 *Naive bayes* rasio pembagian 75:25

Rasio	Classifier	Presisi	Recall	Akurasi	F1-Score
75:25	Gaussian	0.76218	0.7625	0.7197	0.7434

Rasio kelima yang diuji oleh peneliti adalah 80% data dijadikan data latih dan 20% data lainnya menjadi data uji. Melalui pengujian tersebut, didapatkan hasil nilai akurasi sebesar 0.72512, dilanjutkan dengan nilai presisi sebesar 0.76189, lalu nilai recall sebesar 0.74506, dan terakhir nilai F1-score sebesar 0.73076. Hasil ditunjukkan pada Tabel 5.13.

Tabel 5.13 *Naive bayes* rasio pembagian 80:20

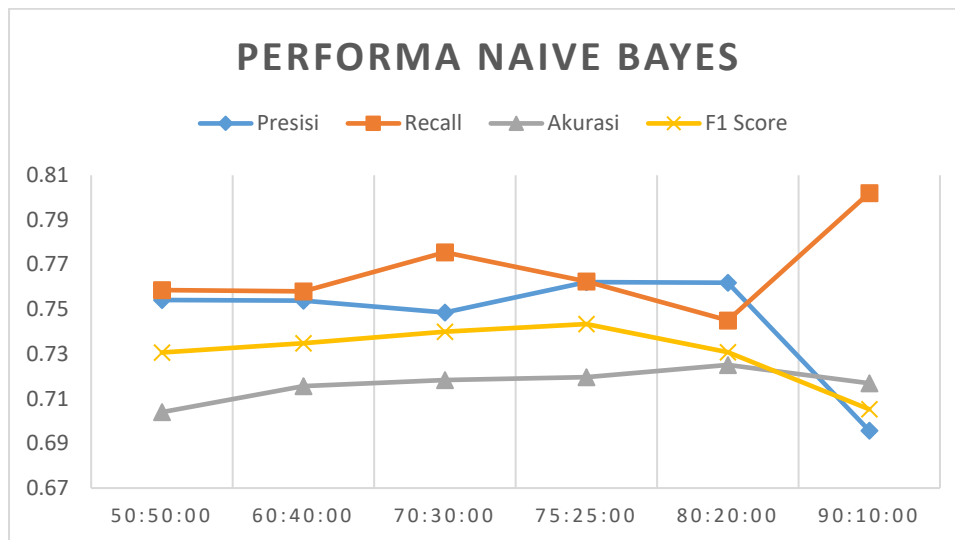
Rasio	Classifier	Presisi	Recall	Akurasi	F1-Score
80:20	Gaussian	0.76189	0.74506	0.72512	0.73076

Rasio keenam yang diuji oleh peneliti adalah 90% data dijadikan data latih dan 10% data lainnya menjadi data uji. Melalui pengujian tersebut, didapatkan hasil nilai akurasi sebesar 0.71698, dilanjutkan dengan nilai presisi sebesar 0.69571, lalu nilai recall sebesar 0.80208, dan terakhir nilai F1-score sebesar 0.70527. Hasil ditunjukkan pada Tabel 5.14.

Tabel 5.14 *Naive bayes* rasio pembagian 90:10

Rasio	Classifier	Presisi	Recall	Akurasi	F1-Score
90:10	Gaussian	0.69571	0.80208	0.71698	0.70527

Hasil pembagian data untuk pemodelan algoritma *naive bayes* ditunjukkan pada Gambar 5.3. Melalui grafik di tersebut dapat dilihat bahwa rasio kelima yaitu 80% data latih dan 20% data uji menghasilkan nilai akurasi tertinggi dan presisi tertinggi. Nilai performa dari rasio ini juga lebih stabil dibandingkan rasio lainnya. Maka dari itu peneliti memutuskan untuk menggunakan rasio 80% data latih dan 20% data uji untuk pemodelan *naive bayes*.

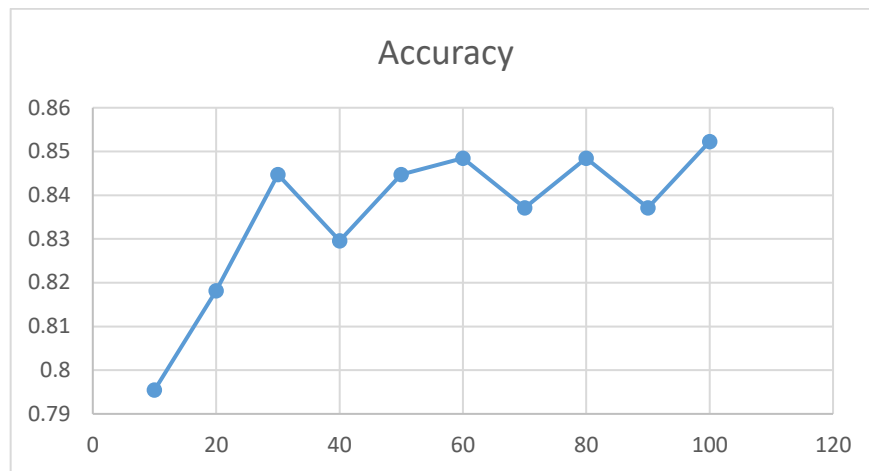


Gambar 5.2 Perbandingan performa *naive bayes*

Berdasarkan hasil dari perbandingan pembagian data yang telah dilakukan, algoritma *random forest* memiliki performa terbaik pada rasio 75:25 (latih:uji). Sedangkan algoritma *naive bayes* memiliki performa terbaik pada rasio 80:20 (latih:uji). Hal ini mengacu pada metode uji *confusion matrix* yang menggunakan Nilai *precision*, *recall*, *accuracy*, dan *F1-Score*, sebagai pengukurannya. Semakin tinggi hasil dari keempat nilai tersebut menunjukkan semakin baik pula model yang telah dibuat. Oleh karena itu pada penelitian ini akan digunakan rasio 75:25 untuk pemodelan algoritma *random forest* dan rasio 80:20 untuk pemodelan algoritma *naive bayes*.

5.1.3. Pemodelan dengan *Random Forest*

Skicit-learn merupakan *library* pada python yang digunakan untuk memodelkan *random forest*. *Library* tersebut berguna untuk membangun model *random forest* untuk melakukan klasifikasi data kualitas air. Langkah pertama yang dilakukan mencari $N_estimators$ atau banyaknya pohon dalam *random forest*. Dalam menentukan $N_estimators$ dilakukan percobaan dengan memasukkan jumlah $N_estimators$ dengan kelipatan 10 dimulai dari 10 sampai 100. Hasil pencarian $N_estimators$ dapat ditunjukkan pada Gambar 5.3.

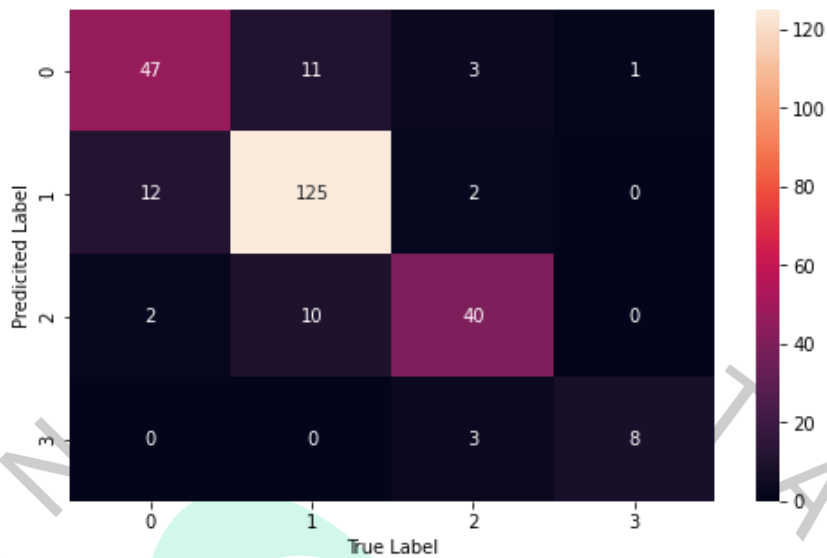


Gambar 5.3 $n_estimators$

Pada grafik diatas, jumlah $n_estimators$ yang memiliki akurasi paling tinggi merupakan $n_estimators$ yang akan dipilih. Dengan begitu model akan menghasilkan performa terbaik saat diterapkan pada metode uji *confusion matrix*. Selanjutnya pemodelan dibuat menggunakan 100 sebagai $n_estimators$. Hasil uji dari pemodelan ditunjukkan pada Tabel 5.15 dan Gambar 5.4.

Tabel 5.15 Hasil pemodelan *random forest*

Rasio	$n_estimators$	Precision	Recall	Accuracy	F1 Score	Waktu Komputasi
75:25	100	0.88112	0.79891	0.85227	0.83351	0.00007



Gambar 5.4 Hasil klasifikasi algoritma *random forest*

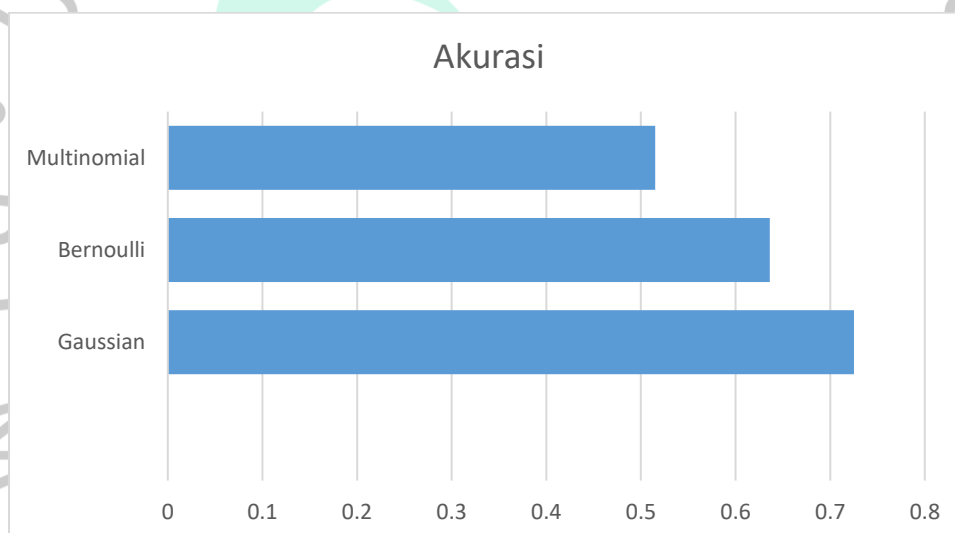
Hasil *precision*, *recall*, *accuracy*, dan *F1-Score* yang didapatkan dengan *n_estimators* tersebut mencapai hasil yang baik. Karena hasil *accuracy* yang diberikan melebihi angka 0.7, dan nilai *precision* berada di atas angka 0.8, lalu *recall* juga tidak kurang dari nilai *accuracy* sehingga menghasilkan nilai *F1-score* diatas 0.8. Perolehan keempat nilai ini baik dan juga stabil karena tidak ada nilai yang berada di bawah 0.7. Hasil klasifikasi kualitas air DKI Jakarta menggunakan algoritma *random forest* ditampilkan pada Tabel 5.16.

Tabel 5.16 Hasil klasifikasi kualitas air DKI Jakarta *random forest*

No.	Keke- ruhan	Air Raksa	Besi (Fe)	Fluo- rida	Kesadahan	Krom Heksavalen	Mangan (Mn)	Nitrat	Nitrit	Seng (Zn)	Sulfat	Detergen	KMnO4	Total Coliform	E.Coli	Kategori Prediksi
1	0.55	0	0.08	0.05	181.53	0	0.05	3.94	0.006	0.04	18.27	0.05	1.42	100	0	Tercemar Ringan
2	17.44	0.0003	0.08	0.05	117	0	0	0.67	0.027	0	12.75	0	6	4.4	0	Tercemar Ringan
3	0.37	0.3	0.8	0.8	90.44	0.3	0.5	3.15	0.3	0.3	16.24	0.5	2.5	0	0	Memenuhi Baku Mutu
4	0.4	0	0.08	0.21	273.95	0	0.14	4.92	0.008	0.04	41.28	0.07	2.69	200	0	Tercemar Ringan
5	5.02	0.0003	0.08	0	88	0	3	0.31	0.009	0.01	28.02	0	3	78	4.2	Tercemar Ringan
6	0	0.0003	0	0	93	0.003	0.05	2	0	0	9	0.03	0	0	0	Memenuhi Baku Mutu
7	1	0.0003	0	0	59	0.003	0	0	0	0.01	7	0.03	0	0	0	Memenuhi Baku Mutu
8	16.74	0.3	0.13	1.19	219.61	0.3	1.32	0.8	0.2	0.1	7	0.8	26.77	6100	100	Tercemar Sedang
9	0.52	0	0.08	0.05	129.4	0	0.05	3.42	0.007	0.01	1.73	0.05	0.5	0	0	Memenuhi Baku Mutu
10	2	0.0003	0	0	84	0.003	0.05	0	0	0.01	8	0.03	0	0	0	Memenuhi Baku Mutu

5.1.4. Pemodelan dengan *Naïve Bayes*

Skicit-learn merupakan *library* pada *python* yang digunakan untuk memodelkan *naïve bayes*. *Library* tersebut berguna untuk membangun model *naïve bayes* untuk melakukan klasifikasi data kualitas air. Pada penelitian ini terdapat 3 jenis *classifier naïve bayes* yang akan dibandingkan yaitu *gaussian*, *bernoulli*, dan *multinomial*. Peneliti melakukan perbandingan hasil dari ketiga jenis *classifier naïve bayes* tersebut dan menentukan jenis *classifier naïve bayes* yang paling sesuai dengan data yang digunakan pada penelitian ini. Hasil dari perbandingan ketiga jenis *naïve bayes* dapat ditunjukkan pada Gambar 5.5.

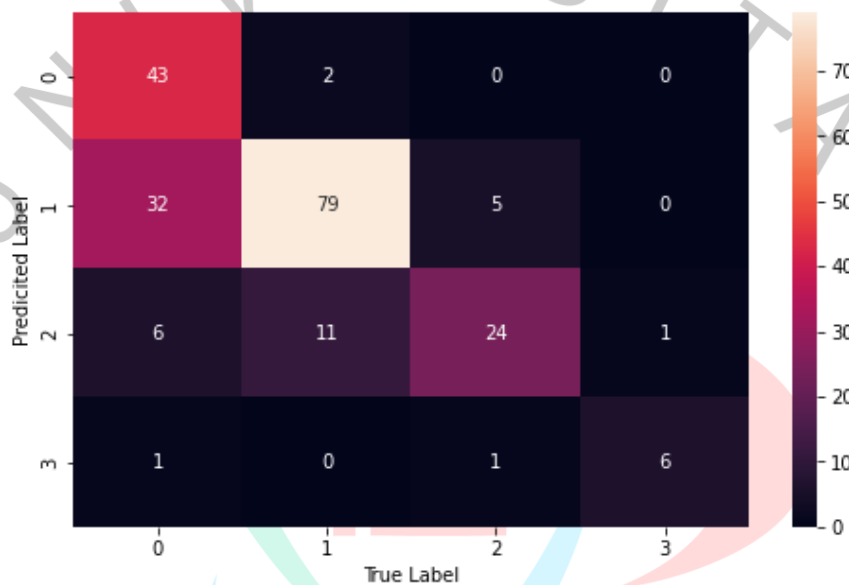


Gambar 5.5 Perbandingan jenis *classifier naïve bayes*

Berdasarkan tabel perbandingan diatas, nilai *precision*, *recall*, *accuracy*, dan *F1-score* paling tinggi dimiliki oleh *gaussian classifier*. Nilai dari *classifier* lainnya jauh lebih rendah dan tidak stabil. Oleh karena itu, *naïve bayes* dengan jenis *classifier gaussian* diterapkan dalam pemodelan *naïve bayes*. Hasil uji pemodelan menggunakan algoritma *naïve bayes* ditunjukkan pada Tabel 5.17 dan Gambar 5.6

Tabel 5.17 Hasil pemodelan *naive bayes*

Rasio	classifier	Precision	Recall	Accuracy	F1 Score	Waktu Komputasi
75:25	Gaussian	0.76712	0.76204	0.72348	0.7451	0.00005



Gambar 5.6 Hasil klasifikasi algoritma *naive bayes*

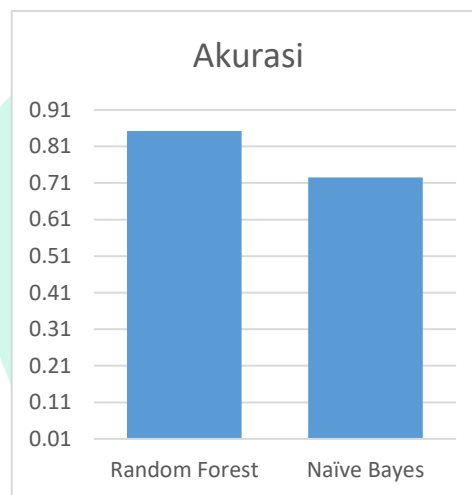
Hasil *precision*, *recall*, *accuracy*, dan *F1-Score* yang didapatkan dengan *gaussian classifier* tersebut mencapai hasil yang baik. Karena semua performa melebihi angka 0.7 Perolehan keempat nilai ini baik dan juga stabil karena tidak ada nilai yang berada di bawah 0.7. Hasil klasifikasi kualitas air DKI Jakarta menggunakan algoritma *naive bayes* ditampilkan pada Tabel 5.18.

Tabel 5.18 Hasil klasifikasi kualitas air DKI Jakarta *naive bayes*

No.	Keke- ruhan	Air Raksa	Besi (Fe)	Fluo- rida	Kesa- dahan	Krom Heksavalen	Mangan (Mn)	Nitrat	Nitrit	Seng (Zn)	Sulfat	Detergen	KMnO4	Total Coliform	E.Coli	Kategori Prediksi
1	0.55	0	0.08	0.05	181.53	0	0.05	3.94	0.006	0.04	18.27	0.05	1.42	100	0	Tercemar Ringan
2	17.44	0.0003	0.08	0.05	117	0	0	0.67	0.027	0	12.75	0	6	4.4	0	Tercemar Sedang
3	0.37	0.3	0.8	0.8	90.44	0.3	0.5	3.15	0.3	0.3	16.24	0.5	2.5	0	0	Memenuhi Baku Mutu
4	0.4	0	0.08	0.21	273.95	0	0.14	4.92	0.008	0.04	41.28	0.07	2.69	200	0	Tercemar Ringan
5	5.02	0.0003	0.08	0	88	0	3	0.31	0.009	0.01	28.02	0	3	78	4.2	Tercemar Sedang
6	0	0.0003	0	0	93	0.003	0.05	2	0	0	9	0.03	0	0	0	Memenuhi Baku Mutu
7	1	0.0003	0	0	59	0.003	0	0	0	0.01	7	0.03	0	0	0	Memenuhi Baku Mutu
8	16.74	0.3	0.13	1.19	219.61	0.3	1.32	0.8	0.2	0.1	7	0.8	26.77	6100	100	Tercemar Sedang
9	0.52	0	0.08	0.05	129.4	0	0.05	3.42	0.007	0.01	1.73	0.05	0.5	0	0	Memenuhi Baku Mutu
10	2	0.0003	0	0	84	0.003	0.05	0	0	0.01	8	0.03	0	0	0	Memenuhi Baku Mutu

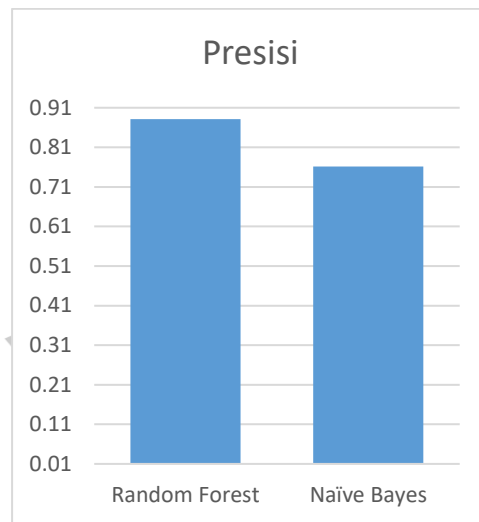
5.2. Hasil dan Pembahasan Pengujian *Confusion Matrix*

Pengujian terhadap hasil klasifikasi yang telah dilakukan berguna untuk mengetahui apakah model telah dibangun dengan baik atau tidak. Serta untuk mengetahui model mana yang memiliki performa lebih baik dengan membandingkan beberapa parameter dari *confusion matrix* yaitu *precision*, *recall*, *accuracy*, dan *F1-score*.



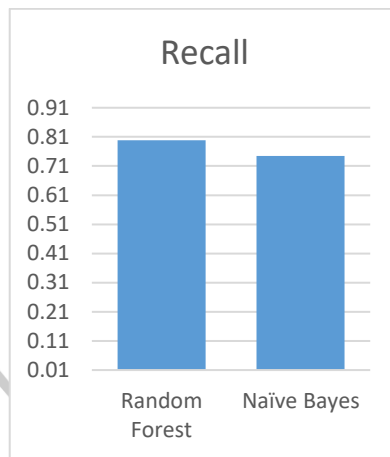
Gambar 5.7 Perbandingan akurasi *random forest* dan *naïve bayes*

Pada hasil akurasi, algoritma *random forest* memiliki nilai lebih tinggi yakni sebesar 0.85227, dibandingkan dengan algoritma *naïve bayes* yang memiliki nilai akurasi sebesar 0.72512. Semakin tinggi nilai akurasi, semakin baik juga model yang telah dibuat. Hal ini karena akurasi merupakan jumlah nilai yang diprediksi benar baik positif maupun negatif dari keseluruhan prediksi. Hasil perbandingan akurasi ini menunjukkan seberapa akurat atau tepat model melakukan klasifikasi. Hasil akurasi tersebut ditunjukkan melalui diagram batang yang ada pada Gambar 5.7.



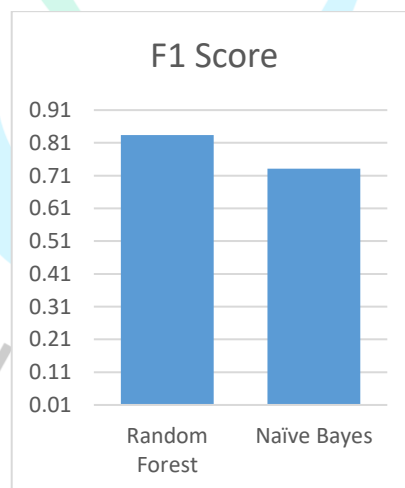
Gambar 5.8 Perbandingan presisi *random forest* dan *naive bayes*

Selanjutnya merupakan perbandingan kedua yakni perbandingan nilai *precision*. Semakin tinggi nilai presisi, maka semakin baik pula model yang dibangun. Karena nilai presisi menunjukkan tingkat keakuratan antara data prediksi dengan data aktual. Nilai presisi didapatkan dari jumlah data diprediksi dengan benar adalah positif, dibagi dengan jumlah keseluruhan data yang diprediksi sebagai positif. Pada penelitian ini, nilai presisi algoritma *random forest* lebih tinggi yakni sebesar 0.88112, sedangkan *naïve bayes* sebesar 0.76189. Hasil presisi ditunjukkan menggunakan diagram batang yang ada pada Gambar 5.8.



Gambar 5.9 Perbandingan recall *random forest* dan *naive bayes*

Pada perbandingan ketiga, parameter recall akan dibandingkan. Semakin tinggi nilai recall yang dimiliki, maka semakin baik model tersebut. Karena nilai recall menggambarkan keberhasilan model saat menemukan kembali sebuah informasi. Nilai recall didapatkan dari pembagian antara jumlah data yang diprediksi positif dengan benar, dan jumlah positif pada data aktual. Hasil recall ditunjukkan menggunakan diagram batang yang ada pada Gambar 5.9.

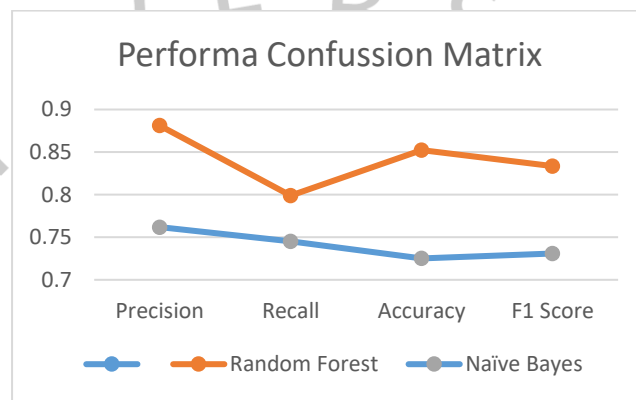


Gambar 5.10 Perbandingan F1-score *random forest* dan *naive bayes*

Perbandingan keempat adalah nilai F1-Score yang merupakan hasil pembobotan dari perbandingan rata-rata precision dan recall. Semakin tinggi nilai

F1-score, maka semakin baik pula model yang telah dibuat. Hasil perbandingan ditunjukkan menggunakan diagram batang yang ada pada Gambar 5.10.

Hasil perbandingan performa *confusion matrix* ditunjukkan melalui diagram garis yang ada pada Gambar 5.11. Hasil tersebut menunjukkan pada bahwa *random forest* mengungguli tiap aspek performa dari pengujian *confusion matrix*.



Gambar 5.11 Perbandingan performa *confussion matrix*