Volume 4 No 2 September 2024
e-ISSN 2776-9690
DOI: http://dx.doi.org/10.53514/jco.v4i2.535

Analisis Perbandingan Metode Naïve Bayes dan K-NN dalam Penentuan Lokasi Layanan Administrasi BPJS Kesehatan di Provinsi Maluku.

Andi Muhammad Irfan[1], Kusrini[1],

[1]Pascasarjana, Magister Teknik Informatika, Universitas Amikom Yogyakarta, Indonesia

*Corresponding author email:
am.irfan@students.amikom.ac.id
kusrini@amikom.ac.id

ABSTRAK

Untuk mencapai Universal Health Coverage (UHC) dimana 98% Penduduk telah memiliki Jaminan Kesehatan Nasional (JKN), menjadi tantangan tersendiri bagi daerah yang geografis dan jarak antar desa yang sangat jauh tentunya terkendala pemerataan pelayanan administrasi BPJS Kesehatan, BPJS Kesehatan telah membuat inovasi Program BPJS Keliling dan BPJS Online, Permasalahannya tidak semua desa dapat dilayani dengan BPJS Keliling dan BPJS Online, sehingga perlu dilakukan pemetaan terhadap desa-desa yang memenuhi syarat untuk dapat dilaksanakan BPJS Keliling dan BPJS Online, Penelitian ini menjelaskan teknik machine learning, khususnya algoritma K-NN dan Naïve Bayes, untuk mensupport pengambilan keputusan dengan pemilihan desa-desa mana yang cocok, layak dan potensial sehingga menghasilkan pelayanan publik yang efektif dan efisien. Hasil percobaan menunjukkan bahwa kedua metode tersebut memiliki tingkat akurasi yang cukup baik, Naïve Bayes mampu mengklasifikasikan desa-desa dengan tingkat akurasi mencapai 94,33%. Performa K-NN tertinggi adalah data yang telah di normalisasi menggunakan Min-Max Scaler dengan akurasi sebesar 95,55%, nilai persisi sebesar 94,27%, nilai recall 95,90% Dan nilai f1 score 94,91%. Oleh karena itu, peneliti merekomendasikan penggunaan algoritma K-NN dengan persyaratan untuk melakukan normalisasi data menggunakan Min-Max Scaler terlebih dahulu, dan nilai k yang optimal adalah 8 untuk menentukan lokasi yang layak mendapatkan layanan Administrasi BPJS Kesehatan.

***Kata Kunci**: Layanan BPJS Kesehatan Online, BPJS Kesehatan Keliling, Naïve Bayes, K-NN*

I. Pendahuluan

Sistem Jaminan Kesehatan Nasional (JKN) di Indonesia telah berproses menuju *Universal Health Coverage (UHC)* atau singkatnya disebut Cakupan Semesta, mencapai target cakupan sebesar 98% dari total penduduk menjadi sangat penting karena berkontribusi pada peningkatan kesejahteraan masyarakat, perlindungan finansial, kesehatan masyarakat yang lebih baik, pertumbuhan ekonomi, dan terciptanya keadilan sosial (Herawati et al., 2020), (Saputro & Fathiyah, 2022). Badan Penyelenggara Jaminan Sosial Kesehatan (BPJS Kesehatan) adalah Badan Hukum Publik yang dibentuk oleh pemerintah untuk menyelenggarakan program jaminan kesehatan.

BPJS Kesehatan sangatlah membantu masyarakat khususnya masyarakat menengah kebawah untuk dapat mengakses layanan kesehatan yang layak. Layanan BPJS Kesehatan sudah menjangkau hampir seluruh wilayah di Indonesia yang tersebar di 127 Kantor Cabang dan 388 Kantor Kabupaten/Kota. Jumlah peserta BPJS Kesehatan hampir setara dengan 97% jumlah penduduk Indonesia yaitu berjumlah 273 juta jiwa per Juni 2024 (Rizaty, 2024). Namun, banyak diantara peserta BPJS yang tinggal jauh dari pusat layanan sehingga pelayanan disana masih kurang maksimal.

Provinsi Maluku merupakan salah satu provinsi di kawasan Timur Indonesia. Provinsi Maluku merupakan sebuah provinsi kepulauan yang merupakan gugus pulau-pulau kecil yang berjumlah 1.392 pulau. Provinsi Maluku yang juga dikenal sebagai 'Provinsi Seribu Pulau'. Secara administratif, provinsi Maluku terbagi atas 9 kabupaten dan 2 kota dengan jumlah 118 kecamatan dan 1.240 desa dan kelurahan. Ciri geografis dan topografi yang khas di Maluku ini memiliki dampak yang signifikan pada perkembangan capaian layanan administrasi pendaftaran BPJS Kesehatan. Oleh karena itu BPJS Kesehatan, terutama di Provinsi Maluku perlu mengupayakan upaya ekstra untuk mencapai target *UHC*. Untuk mencapai targetnya BPJS Kesehatan telah membuat Inovasi layanan BPJS Keliling dan BPJS

 DOI: http://dx.doi.org/10.53514/jco.v4i2.535

Online, layanan ini bertujuan mendekatkan layanan BPJS Kesehatan ke masyarakat dengan melayani masyarakat dengan mobil BPJS Keliling dan melayani masyarakat secara online di Desa. Meskipun demikian, muncul masalah yang perlu diatasi, yaitu bahwa tidak semua desa dapat dilayani karena terbatasnya SDM, Waktu dan Anggaran. Oleh karena itu, pemilihan desa-desa yang layak dilayani harus didasarkan pada prinsip efektif dan efisien . Selama ini, pemilihan desa masih dilakukan secara manual, yang pada beberapa kasus tidak memberikan hasil yang maksimal.

Pada penelitian ini memberikan solusi menggunakan model machine learning dalam proses pemilihan desa yang layak dilayani BPJS Keliling, BPJS Online dan yang tidak layak. Proses pemilihan ini menggunakan model *machine learning* yang teruji lebih akurat, efektif dan efisien dalam mempelajari pola pola dalam data. *Machine learning* juga dapat meminimalisir terjadinya human error. Model machine learning yang digunakan adalah algoritma *K-NearestNeighbour (KNN)* dan *Naïve Bayes*. Kedua algoritma ini sudah banyak digunakan dalam proses analisis data, termasuk data kependudukan (Riadi et al., 2024), kesehatan (Patgiri & Ganguly, 2021), penjualan (Uludağ & Gürsoy, 2020) dan lain-lain. Pada objek penelitian yang sama, *Naïve Bayes* pernah digunakan untuk melalukan analisis sentimen pengguna mobile JKN oleh Sriani et al. (Sriani & Suhardi, 2024) dan sentiment analisis kualitas layanan BPJS . Selain itu, penelitian yang dilakukan oleh Riadi et al. (Riadi et al., 2024) menggunakan KNN dan Naïve Bayes untuk melakukan klasifikasi proses pemilihan desa untuk pemberian prioritas layanan kependudukan. Pada penelitian lain oleh Agusetiana et al. (Agusetiana & Fitrani, 2022), KNN juga pernah digunakan untuk mengklasifikasikan status layanan di perusahaan Telkom.

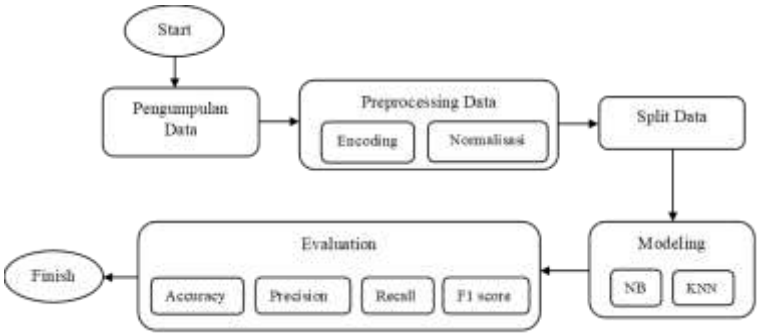
K-Nearest Neighbors (KNN) dan *Naive Bayes* adalah dua algoritma pembelajaran mesin yang digunakan untuk klasifikasi. KNN adalah algoritma non-parametrik yang mengklasifikasikan data berdasarkan kedekatan dengan titik data yang sudah ada; ia mengidentifikasi k sejumlah tetangga terdekat dan menentukan kelas mayoritas dari tetangga tersebut sebagai prediksi. *Naive Bayes*, sebaliknya, adalah algoritma probabilistik yang menggunakan Teorema Bayes dan mengasumsikan independensi antara fitur-fitur; ia menghitung probabilitas dari setiap kelas yang diberikan fitur dan memilih kelas dengan probabilitas tertinggi sebagai prediksi. KNN mudah dipahami dan diimplementasikan tetapi bisa lambat dengan data besar, sementara Naive Bayes cepat dan efektif untuk data tinggi dimensi dengan asumsi independensi yang sering tidak realistis.

Kontibusi dalam penelitian ini adalah untuk memudahkan proses pemilihan desa yang diprioritaskan atau layak dilayani. Proses klasifikasi ini akan menghasilkan klasifikasi desa ke beberapa kategori yaitu yang tidak layak dilayani, layak dilayani BPJS online dan layak dilayani BPJS keliling. Berbagai faktor akan menjadi pertimbangan seperti contohnya presentasi keaktifan peserta, ketersedian jaringan komunikasi data, dan tingkat kesulitan akses dari desa ke kota. Data akan melalui tahap pembersihan hingga masuk ke tahap pengolahan dan evaluasi. Evaluasi akan menggunakan confusion metrik dan nilai akurasi. Hasil akurasi akan menunjukan seberapa akurat model dalam mengklasifikan desa ke kategori yang tepat.

II. Metode Penelitian

A. Tahapan Penelitian

Penelitian ini akan dilakukan secara kuantitatif dengan beberapa langkah yaitu, pengumpulan data, preprocessing data, modeling dan evaluasi. Langkah-langkah penelitian dapat dilihat pada flowchart di Gambar 1.



Gambar 1. Flowchart Penelitian

DOI: http://dx.doi.org/10.53514/jco.v4i2.535

Berdasarkan flowchart pada Gambar 1, penelitian dimulai dengan proses pengumpulan data. Pengumpulan data didapat dari BPJS Kesehatan Kesehatan Cabang Ambon dan Dinas Kependudukan dan Catatan Sipil Provinsi Maluku Setelah data yang dibutuhkan sudah lengkap, data akan melalui proses Preprocessing yaitu data cleansing, encoding dan normalisasi data. Data *Cleansing* yang dilakukan berupa *handling missing value*, yaitu menghapus data-data yang tidak ada value nya. Encoding untuk merubah data kategorikal menjadi data numerikal. Normalisasi data menggunakan 2 macam metode yaitu metode Z-score dan Minmax Scaler. Setelah data bersih dan range data sudah sama, selanjutnya data akan masuk ke tahap modelling atau pengolahan data menggunakan model machine learning. Model *machine learning* yang digunakan adalah KNN dan Naïve Bayes. Setelah model di training oleh data yang masukan, model akan dievaluasi menggunakan confusion metrics dan akurasi.

B. Data

Data terdiri dari 1233 data desa yang ada di provinsi Maluku. Data ini memiliki 8 fitur yaitu Persentasi Cakupan Kepesertaan JKN, Persentasi keaktifan peserta, Persentasi menunggak, Ketersediaan Jaringan di kantor desa, Tingkat kesulitan akses dari desa ke kota, Jarak tempuh dari desa ke kota, Tingkat kelayakan untuk layanan BPJS keliling & BPJS Online, terdiri dari tiga kategori: "Tidak Layak", "Layak BPJS Keliling", "Layak BPJS Online". Dataset dapat dilihat pada Tabel 1 dan untuk keterangan nama kolom dapat dilihat pada Tabel 2.

Tabel 1. Dataset desa penerima layanan BPJS

ID	A1	A2	A3	A4	A5	A6	Z1
1	97,21%	94,16%	0,52%	Tidak tersedia	58	Mudah	Layak BPJS Keliling
2	108,43%	99,85%	0,83%	Tidak tersedia	76	Sangat Sulit	Tidak Layak
3	106,84%	85,63%	3,45%	Tersedia	94	Mudah	Layak BPJS Online
4	78,54%	64,62%	7,38%	Tersedia	112	Mudah	Layak BPJS Keliling
5	102,84%	87,55%	2,64%	Tersedia	130	Sangat Sulit	Tidak Layak

Tabel 2. Keterangan kolom dataset

Atribut	Keterangan
ID	ID
A1	Persentasi Cakupan Kepesertaan JKN
A2	Persentasi keaktifan peserta
A3	Persentasi menunggak
A4	Ketersediaan Jaringan di kantor desa
A5	Tingkat kesulitan akses dari desa ke kota
A6	Jarak tempuh dari desa ke kota
Z1	Tingkat kelayakan untuk layanan BPJS kelling & BPJS Online, terdiri dari tiga kategori: "Tidak Layak", "Layak BPJS Keliling", "Layak BPJS Online"

Data-data ini terangkum dalam data kelayakan dengan delapan atribut. Satu atribut id, satu atribut label, dan enam atribut prediktor. Seperti pada Tabel 1. A1 hingga A3 merupakan atribut prediktor dengan melihat pada presentasi cakupan kepesertaan dan keaktifan peserta BPJS. A4 adalah ketersedian jaringan di kantor desa yang memiliki kategori “Tersedia” dan “Tidak Tersedia”. A5 adalah tingkat kesulitan akses, baik karena akses jalan ataupun topografi dari desa ke kota. Pada atribut ini akan dilakukan proses encoding dengan nilai 1.5 Sangat Sulit, 3.0 Sulit, 4.5 Mudah, 6.0 Mudah Sekali. A6 adalah jarak tempuh dari lokasi desa ke kota tempat pelayanan. Sedangkan Z1 adalah label tingkat kelayakan yang terdiri dari tiga kategori yakni “TIDAK LAYAK”, “LAYAK BPJS ONLINE” dan “LAYAK BPJS KELILING”.

DOI: http://dx.doi.org/10.53514/jco.v4i2.535

C. Preprocessing Data

Preprocessing data terdiri dari 2 tahap yaitu encoding dan normalisasi data. Namun, sebelumnya dimulai dengan pengecekan missing value. Jika terdapat missing value dalam data, maka missing value akan dihapus karena missing value dapat berakibat buruk pada performa model (Purbolaksono et al., 2021).

D. Encoding

Encoding dilakukan pada fitur yang memiliki nilai kategorikal. Nilai kategorikal ini diubah menjadi numerikal dengan label *encoder* yaitu dengan mengubahnya ke skala angka tertentu sesuai tingkatan kategorinya. Kolom yang melalui tahap encoding adalah kolom A6, A4 dan Z1. Pada kolom A6: "Sangat Sulit": 1.5, "Sulit": 3.0, "Mudah": 4.5, "Mudah Sekali": 6.0. Pada kolom A4: "Tersedia": 1, "Tidak tersedia": 0. Pada kolom Z1: "Tidak Layak": 0, "Layak BPJS Keliling": 1, "Layak BPJS Online": 2.

E. Normalisasi Data

Normalisasi data bertujuan untuk mengubah nilai atribut data menjadi rentang yang sama, sehingga atribut data dengan skala yang berbeda dapat dibandingkan secara adil. Normalisasi data menggunakan Z-score dan Min-Max Scaler.

F. Z-score

Z-score, mengubah data sehingga memiliki rata-rata (mean) 0 dan standar deviasi (standard deviation) 1. Metode ini digunakan untuk membuat data memiliki skala yang seragam, terutama ketika data memiliki outlier atau rentang nilai yang sangat berbeda. Rumus Z-score adalah:

$$Z = \frac{(X-\mu)}{\sigma} \quad (1)$$

di mana:

X adalah nilai data.

μ adalah rata-rata dari data.

σ adalah standar deviasi dari data.

G. Min-Max Scaler

Min-Max Scaler mengubah data ke dalam rentang [0, 1] atau rentang lain yang ditentukan pengguna. Metode ini digunakan ketika kita ingin menjaga distribusi asli data tetapi menyesuaikan skala nilai. Rumus Min-Max Scaler adalah:

$$X' = \frac{(X-X_{min})}{(X_{max}-X_{min})} \quad (2)$$

di mana:

X adalah nilai data asli.

X' adalah nilai data yang telah diskalakan.

Xmin adalah nilai minimum dari data.

Xmax adalah nilai maksimum dari data.

H. Split Data

Setiap dataset dibagi menjadi dua bagian dengan proporsi 80% sebagai data latih dan 20% sebagai data uji. Ini berarti bahwa dari total 1233 record, 986 record digunakan untuk melatih model, dan 247 record digunakan untuk menguji performa model. Pembagian ini bertujuan untuk memastikan bahwa model memiliki cukup data untuk belajar dan membangun pola (data latih), sementara juga memiliki data yang cukup untuk mengevaluasi kinerja dan generalisasi model terhadap data baru (data uji).

I. Modeling

1) KNN

 DOI: http://dx.doi.org/10.53514/jco.v4i2.535

K-Nearest Neighbors (KNN) adalah algoritma non-parametrik yang mengklasifikasikan data berdasarkan kedekatan dengan titik data yang sudah ada (Bablani et al., 2018). Proses pengolahan data KNN dimulai dengan normalisasi data untuk memastikan semua fitur memiliki skala yang sama, lalu memilih nilai k (jumlah tetangga terdekat) yang akan digunakan dalam klasifikasi (Putra et al., 2023). Untuk setiap data yang akan diklasifikasikan, algoritma ini menghitung jarak antara data tersebut dengan semua data dalam dataset menggunakan metrik seperti Euclidean atau Manhattan. Rumus jarak Euclidean antara dua titik data x dan *y* adalah:

$$d(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \quad (3)$$

Setelah itu, algoritma menemukan k data dalam dataset yang memiliki jarak terdekat dengan data yang akan diklasifikasikan. Kelas mayoritas dari k tetangga terdekat ini menjadi prediksi untuk data baru dalam masalah klasifikasi, atau rata-rata nilai dari k tetangga terdekat digunakan dalam regresi.

2) Naïve Bayes

Naive Bayes adalah algoritma probabilistik yang menggunakan Teorema Bayes dan mengasumsikan independensi antara fitur-fitur (Zakaria et al., 2023). Dalam proses pengolahan data, algoritma ini menghitung probabilitas a priori dari setiap kelas P(C) berdasarkan data pelatihan, kemudian menghitung probabilitas bersyarat dari setiap fitur yang diberikan kelas tertentu $P(X_i|C)$. Probabilitas gabungan dari fitur $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ adalah produk dari probabilitas individu mereka yang diberikan kelas:

$$P(\mathbf{X}|C) = \prod_{i=1}^{n} P(X_i|C) \quad (4)$$

Untuk data baru, Naive Bayes menghitung probabilitas posterior untuk setiap kelas menggunakan Teorema Bayes:

$$P(C|\mathbf{X}) = \frac{P(\mathbf{X}|C) \cdot P(C)}{P(\mathbf{X})} \quad (5)$$

dan memilih kelas dengan probabilitas posterior tertinggi sebagai prediksi. KNN mudah dipahami dan diimplementasikan, tetapi bisa lambat dengan data besar, sementara Naive Bayes cepat dan efektif untuk data tinggi dimensi dengan asumsi independensi fitur yang sering tidak realistis.

J. Evaluasi

Evaluasi menggunakan metode *confusion matrix* dan akurasi. *Confusion matrix* akan menghitung nilai TP, FP, TN dan FN. True Positive (TP), ketika model dengan benar mengklasifikasikan sampel positif; True Negative (TN), ketika model dengan benar mengklasifikasikan sampel negatif; False Positive (FP), ketika model salah mengklasifikasikan sampel negatif sebagai positif; dan False Negative (FN), ketika model salah mengklasifikasikan sampel positif sebagai negatif. Tabel Confusion Matrix multi-class dapat dilihat pada Tabel 3.

Tabel 3. Tabel Confusion Matrix multi-class

		Aktual			**Class Precision**
Prediksi	Kelas	A	B	C	
	A	TP_A	E_{BA}	E_{CA}	
	B	E_{AB}	TP_B	E_{CB}	
	C	E_{AC}	E_{BC}	TP_C	
Class Recall					

True Positive (TP): TP_A, TN_B, TN_C
True Negatif (TN):
$TN_A=TP_B+E_{CB}+E_{BC}+TP_C$,
$TN_B=TP_A+E_{CA}+E_{AC}+TP_C$,
$TN_C=TP_A+E_{BA}+E_{AB}+TP_B$,
False Positive (FP): $FP_A= E_{BA} + E_{CA}$, $FP_B= E_{AB} + E_{CB}$, dan $FP_C= E_{AC} + E_{BC}$
False Negative (FN): $FN_A = E_{AB} + E_{AC}$, $FN_B= E_{BA} + E_{BC}$, dan $FN_C= E_{CA} + E_{CB}$

 DOI: http://dx.doi.org/10.53514/jco.v4i2.535

Akurasi adalah metrik yang mengukur seberapa sering model membuat prediksi yang benar dan dihitung sebagai jumlah prediksi benar (TP + TN) dibagi dengan total jumlah prediksi (TP + TN + FP + FN).

III. HASIL DAN PEMBAHASAN

Layanan BPJS Kesehatan Keliling dan BPJS Online dalam rangka meningkatkan efektivitas pelayanan BPJS telah dilaksanakan oleh BPJS Kesehatan, namun kesenjangan dalam layanan desa akibat teknik pemilihan lokasi yang kurang optimal juga merupakan kendala tersendiri. Teknik pemilihan lokasi yang lebih akurat diperlukan untuk mengatasi masalah ini. Penelitian ini membahas penggunaan *machine learning*, khususnya algoritma *Naïve Bayes* dan *K-NN*, untuk menentukan lokasi pelayanan administrasi BPJS Kesehatan dan membandingkan kinerja keduanya untuk menemukan algoritma yang optimal, akurat dan stabil. Hasilnya akan memberikan rekomendasi algoritma yang tepat untuk pemilihan lokasi layanan administrasi BPJS Kesehatan.

Pada tahapan ini dilakukan normalisasi pada fitur-fitur prediktor, normalisasi membantu menyamakan skala semua fitur sehingga fitur data dengan skala yang berbeda dapat dibandingkan secara adil. Z-score mengubah data sehingga memiliki mean 0 dan standar deviasi 1. Sedangkan Min-Max Scaler mengubah data sehingga memiliki nilai minimum dan maksimum menjadi 0 dan 1.

Tabel 4. Hasil normalisasi Z-score

ID	A1	A2	A3	A4	A5	A6
0	-0.343	0.406	-0.618	-3.796	-1.730	0.544
1	0.248	0.739	-0.516	-3.796	-1.728	-1.361
2	0.165	-0.094	0.349	0.263	-1.725	0.544
3	-1.325	-1.325	1.645	0.263	-1.722	0.544
4	-0.046	0.018	0.080	0.263	-1.719	-1.361
...	...	...	...	...	...	...
1232	-0.337	-0.336	-0.274	0.263	1.730	0.544

Tabel 5. Hasil normalisasi Min-max scaler

ID	A1	A2	A3	A4	A5	A6
0	0.124	0.216	0.016	0.0	0.000	0.667
1	0.163	0.238	0.026	0.0	0.0008	0.000
2	0.157	0.184	0.106	1.0	0.0016	0.667
3	0.061	0.105	0.227	1.0	0.0024	0.667
4	0.144	0.191	0.081	1.0	0.0032	0.000
...	...	...	...	...	...	...
1232	0.125	0.168	0.048	1.0	1.0	0.667

Tabel 6. Perbaningan Statistik data hasil normalisasi

ATRIBUT	JENIS DATASET	NILAI MINIMUM	NILAI MAKSIMUM	NILAI RATA-RATA	NILAI STD
A1	Data Asli	0,60	3,56	1,04	0,19
	Z-Score	-2,28	13,26	9,28	1,00
	Min-Max	0,00	1,00	0,15	0,06
A2	Data Asli	0,37	3,02	0,87	0,17
	Z-Score	-2,95	12,57	-6,57	1,00
	Min-Max	0,00	1,00	1,00	0,06
A3	Data Asli	0,00	0,33	0,02	0,03
	Z-Score	-0,79	9,95	-1,15	1,00
	Min-Max	0,00	1,00	0,07	0,09
A4	Data Asli	0,00	1,00	0,94	0,25
	Z-Score	-3,80	0,26	-1,15	1,00
	Min-Max	0,00	1,00	0,94	0,25
A5	Data Asli	58,00	22234,00	11146,00	6409,45
	Z-Score	-1,73	1,73	0,00	1,00
	Min-Max	0,00	1,00	0,50	0,29
A6	Data Asli	1,50	6,00	3,64	1,57
	Z-Score	-1,36	1,50	-9,09	1,00
	Min-Max	0,00	1,00	0,48	0,35

Dataset asli memperlihatkan variasi yang signifikan dalam nilai minimum, maksimum, rata-rata, dan standar deviasi di antara atribut-atributnya. Setelah proses normalisasi, nilai-nilai ini menjadi

DOI: http://dx.doi.org/10.53514/jco.v4i2.535

lebih seragam. Melalui metode Z-Score, rata-rata (μ) setiap atribut diatur menjadi 0, sementara standar deviasinya (σ) diatur menjadi 1. Di sisi lain, dengan metode Min-Max, nilai minimum semua atribut diubah menjadi 0 dan nilai maksimum menjadi 1.

A. Modeling

1) Naïve Bayes
   Hasil evaluasi *confusion matrix* dapat dilihat di Tabel 7.

Tabel 7. Confusion Matrix Naïve Bayes

		Aktual								
		Dataset Asli			Dataset Z-Score			Dataset Min-Max		
		T-L	L-K	L-O	T-L	L-K	L-O	T-L	L-K	L-O
Prediksi	T-L	99	1	1	99	1	0	99	1	0
	L-K	0	43	6	0	9	1	0	9	1
	L-O	3	3	91	3	37	97	3	37	97

Dari Tabel 7, terlihat bahwa normalisasi mempengaruhi hasil klasifikasi dari *Naïve Bayes*. Hal ini wajar karena dalam proses *Naïve Bayes*, terutama *Gaussian NB*, sudah ada proses normalisasi. Oleh karena itu, normalisasi tambahan tidak diperlukan dalam algoritma NB. Dari hasil pemodelan dapat disimpulkan bahwa naïve bayes dapat memprediksi dengan benar 99 data kelas TIDAK LAYAK pada semua dataset. Sedangkan pada data kelas LAYAK BPJS KELILING 43 data diprediksi benar pada data asli dan 9 pada dataset Z-Score dan Min-Max. Pada data kelas LAYAK BPJS ONLINE 91 diprediksi benar pada data asli dan 97 pada dataset Z-Score dan Min-Max.

Sementara pada data asli terdapat 14 data diprediksi salah, yaitu pada kelas LAYAK BPJS KELILING: 1 diprediksi TIDAK LAYAK dan 3 diprediksi LAYAK BPJS ONLINE, pada kelas TIDAK LAYAK: 3 diprediksi LAYAK BPJS ONLINE, pada kelas LAYAK BPJS ONLINE: 1 diprediksi TIDAK LAYAK dan 6 diprediksi LAYAK BPJS KELILING.

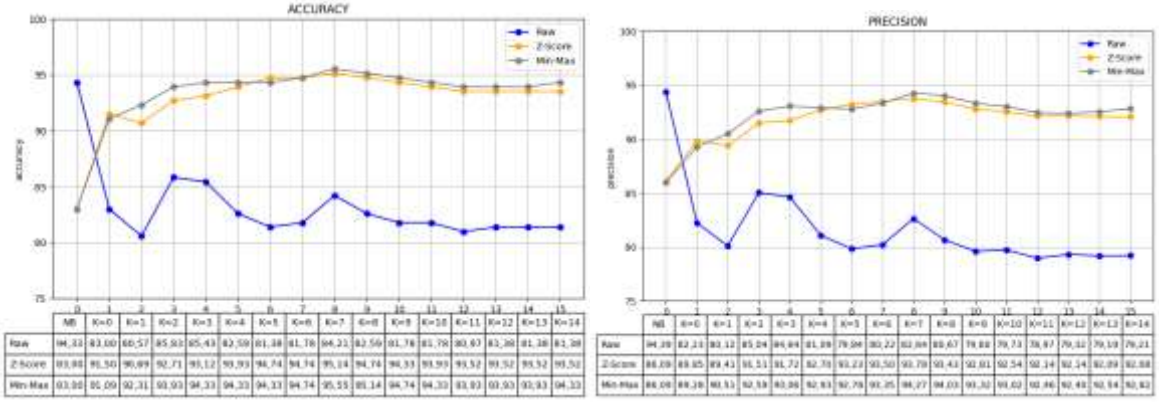
Sedangkan pada dataset Z-Score dan Min-Max terdapat 42 data diprediksi salah, yaitu pada kelas TIDAK LAYAK: 3 diprediksi LAYAK BPJS ONLINE, pada kelas LAYAK BPJS KELILING: 1 diprediksi TIDAK LAYAK dan 37 diprediksi LAYAK BPJS ONLINE, pada kelas LAYAK BPJS ONLINE: 1 diprediksi LAYAK BPJS KELILING.

Tabel 8. Confusion Matrix KNN

			Aktual								
			Dataset Asli			Dataset Z-Score			Dataset Min-Max		
			T-L	L-K	L-O	T-L	L-K	L-O	T-L	L-K	L-O
Prediksi	K=1	T-L	87	2	12	100	0	4	100	0	4
		L-K	3	40	8	0	41	9	0	41	10
		L-O	12	5	78	2	6	85	2	10	84
	K=2	T-L	95	3	20	102	2	8	102	1	4
		L-K	2	42	16	0	43	11	0	44	12
		L-O	5	16	62	0	11	79	0	2	82
	K=3	T-L	95	1	13	99	2	3	100	1	2
		L-K	0	42	10	0	42	7	0	43	7
		L-O	7	4	75	3	3	88	2	3	89
	K=4	T-L	96	2	14	100	1	4	101	1	3
		L-K	0	44	13	0	45	9	0	44	7
		L-O	6	1	71	2	1	85	1	2	88
	K=5	T-L	93	3	12	99	2	2	100	1	1
		L-K	1	37	12	0	44	7	0	43	7
		L-O	8	12	74	3	1	89	2	3	90
	K=...	...	...	...	...	...	...	...	...	...	...
	K=8	T-L	95	2	11	100	1	1	101	1	2
		L-K	0	41	15	0	45	7	0	46	7
		L-O	7	4	72	2	1	90	1	0	89
	K=...	...	...	...	...	...	...	...	...	...	...
	K=15	T-L	93	4	10	99	4	1	99	3	0
		L-K	1	34	14	0	43	8	0	44	8
		L-O	8	9	74	3	0	89	3	0	90

 DOI: http://dx.doi.org/10.53514/jco.v4i2.535

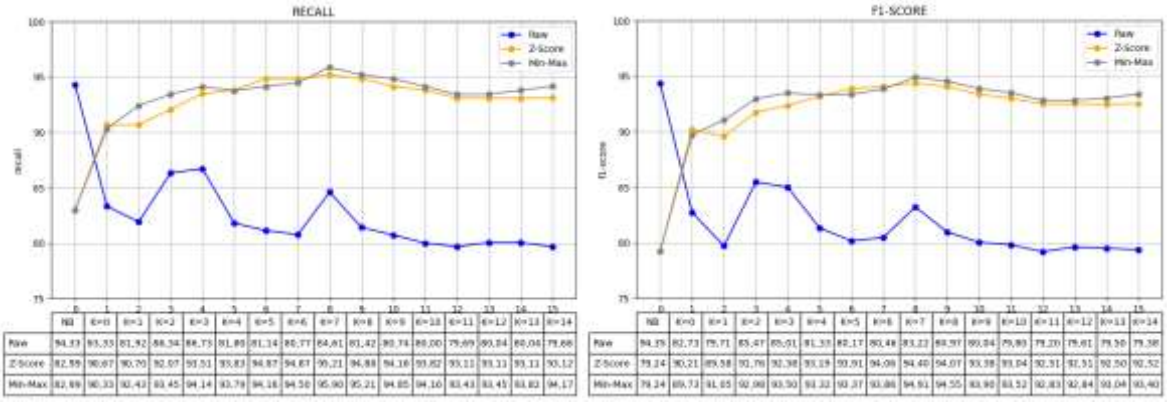
Kami menggunakan alat aplikasi *RapidMiner* untuk memodelkan algoritma K-NN dengan nilai berbeda dari k=1 hingga k=15 dan merangkum hasilnya seperti yang ditunjukkan pada Tabel 8. Hasil pemodelan menunjukkan bahwa pada saat menggunakan dataset ternormalisasi dengan metode min-max scaling, nilai prediksi benar tertinggi adalah k =8, dengan jumlah prediksi yang tepat mencapai 236 dari total 247 data (95,55%). Sebaliknya, nilai prediksi terendah terlihat pada k=2 ketika menggunakan data asli, di mana hanya 199 dari 247 data yang diprediksi dengan benar (80,57%).



Gambar 2. Grafik hasil (a) akurasi, (b) precision

1) Accuracy: Dari analisis confusion matrix diketahui akurasi tertinggi 95,55% dan tercapai ketika menggunakan algoritma K-NN dengan nilai k=8 pada dataset Min-Max. Sedangakan akurasi terendah adalah 80,57% dan terjadi ketika K-NN dengan nilai k=2 digunakan pada data asli. Model Naive Bayes memiliki akurasi sebesar 94,33% pada data asli dan 83,00% pada dataset Z-Score dan Min-Max.

2) Precision: Nilai Presisi Tertinggi adalah 94,27% dan tercapai ketika menggunakan algoritma K-NN dengan nilai k=8 pada dataset Min-Max. Sedangkan Precision terendah adalah 80,12% dan terjadi ketika menggunakan algoritma K-NN dengan nilai k=2 digunakan pada data asli. Algoritma Naive Bayes memiliki precision sebesar 94,39% pada data asli dan 86,09% pada dataset Z-Score dan Min-Max.



Gambar 2 (c) recall, (d) f1-score algoritma KNN

3) *Recall*: Nilai *Recall* (Sensitivitas) tertinggi adalah 95,90% dan tercapai ketika menggunakan algoritma KNN dengan nilai k=8 pada dataset Min-Max. Sedangkan Recall terendah adalah 79,68% dan terjadi ketika menggunakan algoritma K-NN dengan nilai k=15 digunakan pada data asli. Algoritma Naive Bayes memiliki recall sebesar 94,33% pada data asli dan 82,99% pada dataset Z-Score dan Min-Max.

4) *F1-Score* Nilai F1-Score tertinggi adalah 94,91% dan tercapai ketika menggunakan algoritma KNN dengan nilai k=8 pada dataset Min-Max. Sedangkah F1-Score terendah adalah 79,20% dan terjadi ketika menggunakan algoritma K-NN dengan nilai k=12 pada data asli. Algoritma Naive Bayes memiliki F1-Score sebesar 94,35% pada data asli dan 79,24% pada dataset Z-Score dan Min-Max.

IV. Kesimpulan

Layanan BPJS Kesehatan Keliling dan BPJS Online bisa menjadi solusi efektif dan efisien untuk layanan masyarakat khususnya dalam pelayanan administrasi BPJS Kesehatan, namun pemilihan lokasi pelayanan harus dilakukan dengan lebih cermat mengingat mencakup banyak desa yang perlu

 DOI: http://dx.doi.org/10.53514/jco.v4i2.535

dijangkau. Penentuan lokasi desa selama ini seringkali dilakukan secara manual, sehingga menimbulkan ketidakmerataan dalam pelayanan. Dalam mengatasi permasalahan ini, pendekatan klasifikasi dengan menggunakan algoritma *Naïve Bayes* dan *K-NN*, dengan memanfaatkan data kelayakan yang telah ada yang dibagi menjadi data latih dan data uji, terbukti efektif. Hasilnya menunjukkan bahwa algoritma *Naïve Bayes* mampu mengklasifikasikan desa-desa dengan tingkat akurasi mencapai 94,33%. Performa *K-NN* tertinggi adalah data yang telah di normalisasi menggunakan Min-Max Scaler dengan akurasi sebesar 95,55%, nilai persisi sebesar 94,27%, nilai *recall* 95,90% Dan nilai *f1 score* 94,91%, peneliti merekomendasikan penggunaan algoritma K-NN dalam menentukan lokasi layanan administrasi BPJS Kesehatan, dengan persyaratan untuk melakukan normalisasi data menggunakan *Min-Max Scaler* terlebih dahulu, dan nilai k yang optimal adalah 8. Saran untuk penelitian selanjutnya dapat menerapkan metode normalisasi data yang lain atau *hyperparameter tuning* untuk meningkatkan akurasi model.

Daftar Rujukan

Agusetiana, E., & Fitrani, A. S. (2022). Implementasi Data Mining Pada Pelanggan Telkom Menggunakan Metode K-Nearest Neighbor untuk Memprediksi Status Pelayanan. *Prosiding SEMNAS INOTEK (Seminar Nasional Inovasi Teknologi)*, *6*(1), 115–119. https://doi.org/10.29407/inotek.v6i1.2461

Bablani, A., Edla, D. R., & Dodia, S. (2018). Classification of EEG data using k-nearest neighbor approach for concealed information test. *Procedia Computer Science*, *143*, 242–249.

Herawati, H., Franzone, R., & Chrisnahutama, A. (2020). *Universal Health Coverage: Mengukur Capaian Indonesia*.

Patgiri, C., & Ganguly, A. (2021). Adaptive thresholding technique based classification of red blood cell and sickle cell using Naïve Bayes Classifier and K-nearest neighbor classifier. *Biomedical Signal Processing and Control*, *68*, 102745.

Purbolaksono, M. D., Tantowi, M. I., Hidayat, A. I., & Adiwijaya, A. (2021). Perbandingan support vector machine dan modified balanced random forest dalam deteksi pasien penyakit diabetes. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, *5*(2), 393–399.

Putra, R. F., Zebua, R. S. Y., Budiman, B., Rahayu, P. W., Bangsa, M. T. A., Zulfadhilah, M., Choirina, P., Wahyudi, F., & Andiyan, A. (2023). *Data Mining: Algoritma dan Penerapannya*. PT. Sonpedia Publishing Indonesia.

Riadi, I., Yudhana, A., & M. Rosyidi Djou. (2024). Comparative Analysis of Naïve Bayes and K-NN in Determining Location of Mobile Population Services. *Jurnal CoSciTech (Computer Science and Information Technology)*, *4*(3), 733–742. https://doi.org/10.37859/coscitech.v4i3.6543

Rizaty, M. A. (2024). Data Jumlah Peserta BPJS Kesehatan di Indonesia hingga 1 Juni 2024. *Dataindonesia.Id*. https://dataindonesia.id/kesehatan/detail/data-jumlah-peserta-bpjs-kesehatan-di-indonesia-hingga-1-juni-2024

Saputro, C. R. A., & Fathiyah, F. (2022). Universal Health Coverage: Internalisasi Norma di Indonesia. *Jurnal Jaminan Kesehatan Nasional*, *2*(2), 204–216.

Sriani, S., & Suhardi, S. (2024). ANALISIS SENTIMEN PENGGUNA APLIKASI MOBILE JKN MENGGUNAKAN ALGORITMA NAÏVE BAYES CLASSIFIER DAN C4. 5. *JOURNAL OF SCIENCE AND SOCIAL RESEARCH*, *7*(2), 555–563.

Uludağ, O., & Gürsoy, A. (2020). On the financial situation analysis with KNN and naive Bayes classification algorithms. *Journal of the Institute of Science and Technology*, *10*(4), 2881–2888.

Zakaria, P. S., Julianto, R., & Bernada, R. S. (2023). Implementasi Naive Bayes Menggunakan Python Dalam Klasifikasi Data. *Buletin Ilmiah Ilmu Komputer Dan Multimedia (BIIKMA)*, *1*(1), 126–131.