



Analisis Perbandingan Metode Naïve Bayes dan K-NN dalam Penentuan Lokasi Layanan Kependudukan *Mobile*

Imam Riadi¹, Anton Yudhana², M. Rosyidi Djou³

Email: imam.riadi@is.uad.ac.id, eyudhana@ee.uad.ac.id, djou2107048009@webmail.uad.ac.id

¹Program Studi Sistem Informasi, Universitas Ahmad Dahlan, Yogyakarta, Indonesia

²Program Studi Teknik Elektro, Universitas Ahmad Dahlan, Yogyakarta, Indonesia

³Program Studi Informatika, Universitas Ahmad Dahlan, Yogyakarta, Indonesia

Diterima: 30 Desember 2023 | Direvisi: - | Disetujui: 1 Januari 2024

©2020 Program Studi Teknik Informatika Fakultas Ilmu Komputer,
Universitas Muhammadiyah Riau, Indonesia

Abstrak

Tantangan geografis dan jarak antar desa menjadi kendala dalam pemerataan pelayanan kependudukan dan pencatatan sipil. Sehingga perlu adanya campur tangan program jemput bola atau pelayanan mobile. Permasalahannya adalah tidak semua desa dapat dilayani dengan layanan mobile, sehingga perlu dilakukan pemetaan terhadap desa-desa yang memenuhi syarat menjadi lokasi pelayanan mobile. Penelitian ini menjelaskan teknik machine learning, khususnya algoritma K-NN dan Naïve Bayes, untuk mengatasi masalah pemilihan lokasi yang layak. Hasil percobaan menunjukkan bahwa kedua metode tersebut memiliki tingkat akurasi yang cukup baik, dengan K-NN mencapai tingkat akurasi tertinggi yaitu 97,14% pada dataset yang dinormalisasi dengan metode Normalisasi Min-Max (NMM). Sebaliknya, Naïve Bayes menunjukkan nilai akurasi yang tingginya di semua dataset. Oleh karena itu, penelitian ini merekomendasikan penggunaan algoritma K-NN dengan nilai K=1 dan K=2 untuk menentukan lokasi yang layak mendapatkan layanan kependudukan *mobile*.

Kata kunci: perbandingan metode, Naïve bayes, K-nn, layanan kependudukan mobile

Comparative Analysis of Naïve Bayes and K-NN in Determining Location of Mobile Population Services

Abstract

Geographical challenges and distance between villages are obstacles in equalizing population and civil registration services. This requires the intervention of a ball pick-up program or mobile service. The problem is that not all villages can be served by mobile services, so it is necessary to map the villages that qualify to be mobile service locations. This research describes machine learning techniques, specifically the K-NN and Naïve Bayes algorithms, to address the problem of selecting eligible locations. Experimental results show that both methods have fairly good accuracy rates, with K-NN achieving the highest accuracy rate of 97.14% on a dataset normalized by the Normalisasi Min-Max (NMM) method. In contrast, Naïve Bayes showed high accuracy values on all datasets. Therefore, this study recommends the use of the K-NN algorithm with a K=2 value to determine locations that are eligible for mobile population services.

Keywords: method comparison, Naïve bayes, K-nn, mobile civil service

1. PENDAHULUAN

Dokumen kependudukan adalah alat penting dalam memeriksa dan memverifikasi identitas penduduk, yang menjadi dasar untuk berbagai layanan publik seperti pelayanan kesehatan, pendidikan, dan bantuan sosial. Dengan memiliki dokumen kependudukan yang sah, masyarakat dapat dengan mudah mengakses layanan yang mereka perlukan. Selain itu, data pendaftaran penduduk dan

dokumen kependudukan digunakan untuk mengumpulkan informasi statistik tentang populasi, termasuk jumlah penduduk, struktur usia, dan tingkat pendidikan. Informasi ini sangat penting dalam proses perencanaan pembangunan, pengambilan kebijakan publik, dan pengambilan keputusan di berbagai sektor. Yang tidak kalah pentingnya, pendaftaran penduduk, dokumen kependudukan, dan pencatatan sipil memberikan fondasi yang kuat untuk melindungi hak-hak individu dan mendukung pembangunan serta keamanan nasional secara menyeluruh [1].

Direktorat Jenderal Kependudukan dan Pencatatan Sipil (Dirjen Dukcapil) berperan sebagai lembaga utama yang mengawasi, mengelola, dan memastikan kelancaran layanan pencatatan kependudukan. Upaya untuk meningkatkan layanan tersebut telah menghasilkan berbagai inovasi, seperti penggunaan tanda tangan digital [2] yang menggantikan metode konvensional dengan stempel pada dokumen kependudukan. Selain itu, penggunaan kertas HVS 80g [3] sebagai alternatif pengganti security paper memungkinkan pencetakan dokumen sendiri, juga dengan adanya platform mandiri Dukcapil (ADM) memfasilitasi masyarakat dalam mencetak dokumen secara mandiri. Penerapan Identitas Kependudukan Digital (IKD) [4] adalah langkah terbaru dalam upaya penyempurnaan layanan.

Kabupaten Alor, yang terletak di Provinsi Nusa Tenggara Timur, memiliki luas wilayah sebesar 2.864,64 km² yang terbagi menjadi 17 Kecamatan dan 175 desa/kelurahan. Secara geografis, wilayah ini terdiri dari sembilan pulau berpenghuni dan 11 pulau lainnya, yang didominasi oleh pegunungan tinggi yang dikelilingi oleh lembah dan jurang. Lebih dari 63,94% wilayah Kabupaten Alor memiliki kemiringan lebih dari 40° [5]. Ciri topografi yang unik ini memiliki dampak yang signifikan pada kemajuan layanan kependudukan dan kepemilikan dokumen kependudukan. Oleh karena itu, Dinas Kependudukan dan Pencatatan Sipil (Disdukcapil), terutama di Kabupaten Alor, perlu mengupayakan upaya ekstra untuk mencapai target kepemilikan dokumen kependudukan.

Untuk mencapai targetnya, Disdukcapil telah menerapkan layanan mobile sebagai salah satu langkah kebijakannya. Layanan ini bertujuan mendekatkan layanan kependudukan ke masyarakat dengan memindahkan tim layanan ke lokasi tertentu khususnya desa. Meskipun demikian, muncul masalah yang perlu diatasi, yaitu bahwa tidak semua desa dapat dilayani dalam satu tahun anggaran karena terbatasnya waktu dan dana. Oleh karena itu, pemilihan desa-desa yang layak dilayani harus didasarkan pada prinsip-prinsip keadilan dan akuntabilitas. Selama ini, pemilihan desa dilakukan secara manual, yang pada beberapa kasus tidak memberikan hasil yang memuaskan.

Solusi yang ditawarkan adalah memanfaatkan teknik machine learning (ML) dengan metode klasifikasi data mining untuk memproses pemilihan desa, sehingga diharapkan pemilihan desa ini dapat berjalan lebih akurat, adil dan merata.

Banyak peneliti sebelumnya yang telah melakukan penelitian tentang pembelajaran mesin (ML) menggunakan metode klasifikasi, dalam bidang kesehatan telah memberikan kontribusi signifikan dalam mendeteksi berbagai penyakit [6] seperti penyakit menular [7], diabetes [8], kanker payudara [9], penyakit ginjal kronis [10], hepatitis C [11], anomali pada sel darah [12], mendeteksi gangguan spektrum autisme atau ASD pada kelompok usia Balita, Anak, Remaja, dan dewasa [13], dan banyak kondisi medis lainnya. Dengan menggunakan teknik pembelajaran mesin seperti K-Nearest Neighbor (K-NN), Decision Tree, dan Support Vector Machine (SVM), data medis dapat dianalisis secara efisien dan akurat untuk mengidentifikasi pola atau karakteristik khusus yang terkait dengan penyakit tertentu. Hal ini memungkinkan diagnosis dini, prediksi risiko, dan pengambilan keputusan medis yang lebih baik. Penerapan metode pembelajaran mesin dalam bidang kesehatan telah menghasilkan pengembangan alat-alat canggih seperti sistem pengenalan imun buatan (AIRS) dan alat deteksi otomatis [14], yang membantu dokter dan tenaga medis dalam mendiagnosis penyakit dengan tingkat akurasi yang lebih tinggi. Ini tidak hanya mempercepat proses diagnosis, tetapi juga meningkatkan pengobatan dan perawatan pasien, selain itu. Selain mendeteksi penyakit algoritma ML juga digunakan untuk mendeteksi emosi manusia [15], jenis tangisan bayi sehat atau sakit [16] dan mengontrol suhu dalam gudang farmasi [17]

Dalam pertanian, teknik Machine Learning (ML) digunakan untuk mendeteksi penyakit pada tanaman, termasuk daun kacang tanah. Dengan menganalisis pola-pola karakteristik pada daun yang terinfeksi, sistem dapat membedakannya dari daun yang sehat menggunakan algoritma K-NN. Hasilnya, penyakit tanaman dapat dideteksi dengan akurasi tinggi. Ini memungkinkan petani untuk lebih cepat mengidentifikasi penyakit pada tanaman kacang tanah dan mengambil langkah-langkah yang sesuai untuk meningkatkan hasil panen serta mencegah penyebaran penyakit yang lebih lanjut. [18], K-NN untuk penentuan tingkat kematangan kopi arabika [19], Naïve bayes digunakan untuk mendeteksi unsur hara dalam tanah [20], [21] dan menentukan jenis tanaman yang harus ditanam petani dengan melihat cuaca [22].

Dalam dunia perikanan, machine learning khususnya algoritma K-NN dan Naïve Bayes juga digunakan dalam memilih dan membedakan antara ikan segar dan ikan busuk, dengan menganalisa citra mata ikan, kedua algoritma tersebut dapat membedakan ikan segar dan ikan busuk, [23]. K-NN, Naive bayes dan sejumlah algoritma lainnya juga digunakan untuk menentukan penampungan hewan sesuai dengan karakteristik hewannya. Penelitian ini membandingkan hasil kinerja masing-masing algoritma untuk mencari algoritma yang terbaik untuk digunakan dalam penentuan penampungan hewan. dan algoritma yang terpilih adalah K-NN dan c4.5 [24]

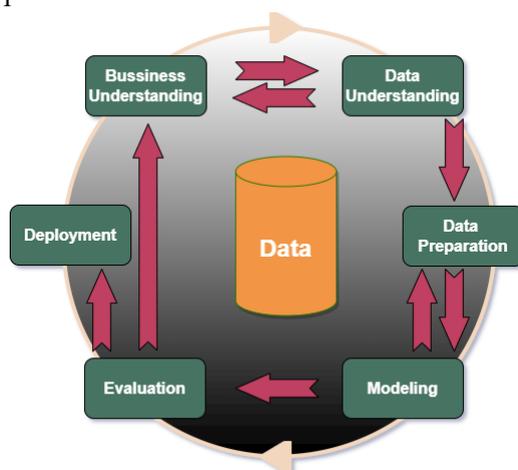
Dari penelitian terdahulu diketahui bahwa metode klasifikasi dengan algoritma Naïve bayes dan K-NN sudah banyak digunakan dalam berbagai bidang pekerjaan, namun masih sangat sedikit yang menggunakan kedua algoritma ini untuk peningkatan pelayanan publik.

Dalam penelitian ini akan dibahas tentang tentang algoritma Naïve Baye dan K-NN dalam penentuan lokasi pelayanan “jemput bola” yang merupakan lanjutan penelitian sebelumnya yang menggunakan algoritma naïve bayes [25] dan K-NN [26] secara terpisah, dengan fokus penelitian pada perbandingan kinerja antara kedua algoritma.

2. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif, yang mana pada akhir penelitian disajikan perbandingan kinerja dari kedua algoritma dalam mengklasifikasikan lokasi pelayanan kependudukan, yang terdiri dari nilai akurasi, presisi, recall, dan F1-Score. Metode evaluasi menggunakan confusion matrix. Confusion matrix adalah alat visualisasi yang biasa digunakan dalam pembelajaran terawasi [27] mengevaluasi model klasifikasi untuk memperkirakan objek yang benar atau salah., confusion matrix digunakan untuk menghitung nilai akurasi, presisi dan recall [8], [28], [29]. Confusion matrix biasanya terdiri dari kolom dan baris yang masing-masing mewakili nilai hasil prediksi dan nilai aktual.

Dalam penelitian ini, penulis mengacu pada serangkaian langkah yang telah menjadi standar dalam proses CRISP-DM. Sebelumnya, telah banyak para peneliti sebelumnya yang mengadopsi kerangka kerja ini dalam konteks proyek data mining [30]–[34], sebagaimana pada Gambar 1



Gambar 1 Tahapan Penelitian CRISP-DM

Adapun tahapan data proyek data mining menggunakan standar CRISP-DM adalah sebagai berikut:

1. *Business Understanding*. Tahap pertama dalam penelitian ini adalah mendefinisikan dengan jelas tujuan dari penelitian ini. Dalam konteks ini, tujuannya adalah mencari algoritma yang cocok untuk mengidentifikasi dan menentukan lokasi layanan kependudukan dan pencatatan sipil *mobile*. Layanan *mobile* ini memiliki dampak besar pada efisiensi pelayanan publik dan aksesibilitas masyarakat terhadap layanan administratif seperti akta kelahiran, KTP, dan Kartu Keluarga
2. *Data Understanding* Dalam tahap ini, data akan digunakan dalam penentuan lokasi layanan kependudukan dan pencatatan sipil *mobile* dikumpulkan, diperiksa, dan dipahami secara komprehensif. Memahami karakteristik data ini adalah langkah penting dalam merencanakan dan merancang pemodelan yang tepat.
3. *Data Preparation*. Data yang telah dikumpulkan harus dipersiapkan sebelum memasuki tahap pemodelan. Tahapan ini melibatkan pemilihan atribut yang relevan untuk analisis, penanganan nilai yang kosong atau hilang, Pada tahapan ini juga dilakukan Normalisasi. Normalisasi adalah proses menyesuaikan skala nilai atribut dari data agar berada dalam rentang tertentu
4. *Modeling*. Tahap pemodelan melibatkan penggunaan algoritma. Pemilihan algoritma yang tepat adalah langkah kunci dalam memastikan keberhasilan model, metode yang digunakan adalah klasifikasi data mining, dengan algoritma naïve bayes dan K-NN, untuk membangun model yang dapat memprediksi lokasi layanan kependudukan dan pencatatan sipil *mobile*. Alat bantu yang digunakan untuk membantu dalam proses ini adalah RapidMiner.
5. *Evaluation*. Evaluasi adalah langkah penting dalam penelitian ini. Dalam tahapan ini melibatkan pengujian kinerja model yang telah dibangun untuk memastikan keakuratannya. Hasil evaluasi akan digunakan untuk membandingkan kinerja antara algoritma naïve bayes dan K-NN. Hasil evaluasi akan menentukan algoritma mana yang paling cocok untuk menentukan lokasi pelayanan kependudukan dan pencatatan sipil *mobile*, yang akan berdampak pada pengambilan keputusan dan efektivitas layanan yang lebih baik
6. *Deployment*. Ini adalah tahap akhir, di mana informasi yang didapat dari hasil data mining diterjemahkan ke dalam bentuk laporan, rekomendasi atau kode program khusus

3. HASIL DAN PEMBAHASAN

3.1 *Business Understanding*

Layanan kependudukan dan pencatatan sipil *mobile* dalam rangka meningkatkan efektivitas pelayanan kependudukan sudah dilakukan oleh disdukcapil, tetapi ketidakmerataan layanan di desa-desa akibat teknik pemilihan lokasi yang belum optimal menjadi hambatan tersendiri. Untuk mengatasi masalah ini, diperlukan teknik pemilihan lokasi yang lebih tepat. Penelitian ini membahas penggunaan machine learning, khususnya algoritma Naïve Bayes dan K-NN, untuk menentukan lokasi pelayanan kependudukan dan pencatatan sipil *mobile* dan membandingkan kinerja keduanya guna menemukan algoritma yang optimal, akurat dan stabil. Hasilnya akan memberikan rekomendasi algoritma yang tepat untuk pemilihan lokasi layanan kependudukan *mobile*

3.2 Data Understanding dan Data Preparation

Sesuai dengan tujuan penelitian ini adalah penentuan lokasi pelayanan dokumen kependudukan *mobile* menggunakan algoritma Naïve bayes dan K-NN. Data yang dikumpulkan adalah data-data yang berkaitan dengan layanan kependudukan dan jarak serta tingkat kesulitan akses.

Tabel 1. Atribut yang digunakan

Atribut	Keterangan
ID	Nomor Id
A01	Persentase kepemilikan akta kelahiran (%)
A02	Persentase kepemilikan KIA (%)
A03	Persentase kepemilikan KTP (%)
A04	Persentase kepemilikan Kartu Keluarga (%)
A05	Tingkat kesulitan akses dari desa ke kota (skala)
A06	Jarak tempuh dari desa ke kota (km)
L01	Tingkat kelayakan untuk layanan kelling, terdiri dari tiga kategori: "Tidak Layak", "Layak", "Sangat Layak"

Data-data ini terangkum dalam data kelayakan dengan delapan atribut. Satu atribut id, satu atribut label, dan enam atribut prediktor. Seperti pada Tabel 1. A01 hingga A04 merupakan atribut prediktor dengan melihat pada presentasi kepemilikan dokumen kependudukan dan pencatatan sipil, A05 adalah tingkat kesulitan akses, baik karena akses jalan ataupun topografi, A06 adalah jarak tempuh dari lokasi desa ke kota tempat pelayanan kependudukan di dinas kependudukan. Sedangkan L01 adalah label tingkat kelayakan yang terdiri dari tiga kategori yakni “TIDAK LAYAK”, “LAYAK” dan “LAYAK SEKALI”. data kelayakan ini terdiri dari 175 record. Potongan dataset sebagaimana terlihat pada Tabel 2. Data yang telah dikumpulkan harus dipersiapkan sebelum memasuki tahap pemodelan. Hal ini melibatkan pemilihan atribut yang relevan untuk analisis, penanganan nilai yang kosong atau hilang, dan normalisasi data untuk memastikan keseragaman. Proses ini membantu menghilangkan anomali dalam data yang dapat memengaruhi hasil akhir.

Tabel 2. Potongan data kelayakan

ID	A01	A02	A03	A04	A05	A06	L01
1	93,38	18,48	86,26	74,94	SANGAT MUDAH	2	TIDAK LAYAK
2	85,60	25,05	69,67	58,86	SANGAT MUDAH	0	TIDAK LAYAK
3	89,23	16,55	83,32	70,12	SANGAT MUDAH	2	TIDAK LAYAK
4	89,78	18,78	80,53	68,40	SANGAT MUDAH	3	TIDAK LAYAK
5	90,46	13,22	79,14	68,51	SANGAT MUDAH	1	TIDAK LAYAK
6	90,35	24,39	76,69	61,50	SANGAT MUDAH	1	TIDAK LAYAK
7	94,14	40,07	87,81	73,01	SANGAT MUDAH	8	TIDAK LAYAK
8	93,12	34,80	84,11	70,02	SANGAT MUDAH	5	TIDAK LAYAK
...
175	90,48	10,58	76,16	66,67	SANGAT JAUH	47	LAYAK

Pada Atribut A05 telah dilakukan pengkodean atau konversi skala dari label atau kata-kata menjadi angka numerik untuk mempermudah perhitungan. Skala ordinal tingkat kesulitan akses seperti pada pada Tabel 3 sehingga menghasilkan data sebagaimana tabel 4.

Tabel 3. Skala ordinal tingkat kesulitan akses ke desa

Skala ordinal	Keterangan
1.5	Sangat Sulit
3.0	Sulit
4.5	Mudah
6.0	Mudah Sekali

Tabel 4. Potongan data asli

ID	A01	A02	A03	A04	A05	A06	L01
1	93,38	18,48	86,26	74,94	6	2	TIDAK LAYAK
2	85,60	25,05	69,67	58,86	6	0	TIDAK LAYAK
3	89,23	16,55	83,32	70,12	6	2	TIDAK LAYAK
4	89,78	18,78	80,53	68,40	6	3	TIDAK LAYAK
5	90,46	13,22	79,14	68,51	6	1	TIDAK LAYAK
6	90,35	24,39	76,69	61,50	6	1	TIDAK LAYAK
...
175	90,48	10,58	76,16	66,67	1,5	47	LAYAK

Pada tahapan ini juga dilakukan normalisasi pada atribut prediktor, normalisasi membantu menyamakan skala semua atribut sehingga diharapkan memberikan kontribusi yang seimbang [35], [36], sebagai contoh menjadikan skala semua atribut antara 0 dan 1.

Pada penelitian ini normalisasi yang digunakan adalah Normalisasi Z-Score (NZS) dan Normalisasi Min-Max (NMM). NZS menggunakan rumus seperti pada persamaan (2) dan NMM menggunakan rumus dengan persamaan (2) [36]

$$x'_{i,n} = \frac{x_{i,n} - \mu_i}{\sigma_i} \tag{1}$$

Dimana $x'_{i,n}$ adalah data atribut yang telah dinormalisasi, $x_{i,n}$ adalah data atribut ke- i dari record ke- n , μ_i sebagai nilai rata-rata dan σ_i sebagai nilai standar deviasi dari atribut ke- i sebelum normalisasi

$$x'_{i,n} = \frac{x_{i,n} - \min(x_i)}{\max(x_i) - \min(x_i)} (nMax - nMin) + nMin \tag{2}$$

Data $x'_{i,n}$ adalah hasil normalisasi atribut ke- i pada record ke- n , sementara $x_{i,n}$ adalah data asli sebelum normalisasi. $\min(x_i)$ dan $\max(x_i)$ mewakili nilai minimum dan maksimum atribut ke- i dalam dataset, sementara $nMax$ dan $nMin$ adalah inisialisasi skala terbesar dan terkecil yang diinginkan. Potongan dataset asli, dataset NZS dan NMM seperti terlihat pada

Tabel 5

Tabel 5. Data asli, NZS dan NMM

ID	DATASET ASLI						DATASET NZS						DATASET NMM						
	A01	A02	A03	A04	A05	A06	A01	A02	A03	A04	A05	A06	A01	A02	A03	A04	A05	A06	
1	93,38	18,48	86,26	74,94	6	2	0,88	0,68	1,11	1,75	-2,10	-1,41	0,92	0,27	0,90	0,91	0,00	0,02	
2	85,60	25,05	69,67	58,86	6	0	-0,19	1,27	-1,72	0,15	-2,10	-1,49	0,74	0,37	0,41	0,59	0,00	0,00	
3	89,23	16,55	83,32	70,12	6	2	0,31	0,50	0,60	1,27	-2,10	-1,41	0,82	0,24	0,81	0,81	0,00	0,02	
4	89,78	18,78	80,53	68,40	6	3	0,38	0,70	0,13	1,10	-2,10	-1,37	0,83	0,27	0,73	0,78	0,00	0,03	
5	90,46	13,22	79,14	68,51	6	1	0,48	0,20	-0,11	1,11	-2,10	-1,47	0,85	0,19	0,69	0,78	0,00	0,00	
6	90,35	24,39	76,69	61,50	6	1	0,46	1,21	-0,53	0,41	-2,10	-1,45	0,85	0,36	0,62	0,64	0,00	0,01	
7	94,14	40,07	87,81	73,01	6	8	0,98	2,63	1,37	1,56	-2,10	-1,23	0,94	0,59	0,95	0,87	0,00	0,06	
8	93,12	34,80	84,11	70,02	6	5	0,84	2,15	0,74	1,26	-2,10	-1,31	0,94	0,59	0,95	0,87	0,00	0,06	
...
175	90,48	10,58	76,16	66,67	1,5	47	0,48	-0,04	-0,61	0,93	1,26	0,13	0,85	0,15	0,60	0,74	1,00	0,36	

Untuk memberikan pemahaman yang lebih baik tentang efek normalisasi, pada tabel

Tabel 6 disajikan data statistik perbandingan dari ketiga dataset ini.

Tabel 6. Perbandingan Statistik Dataset asli, dataset NZS dan NMM

ATRIBUT	JENIS DATASET	NILAI MINIMUM	NILAI MAKSIMUM	NILAI RATA-RATA	NILAI STDEV
A01	Data asli	54,011	96,875	86,979	7,312
	NZS	-4,509	1,353	0,000	1,000
	NMM	0,000	1,000	0,769	0,171
A02	Data asli	0,000	68,475	10,976	11,074
	NZS	-0,991	5,192	0,000	1,000
	NMM	0,000	1,000	0,160	0,162
A03	Data asli	55,723	89,683	79,774	5,869
	NZS	-4,098	1,688	0,000	1,000
	NMM	0,000	1,000	0,708	0,173
A04	Data asli	28,796	79,643	57,355	10,039
	NZS	-2,845	2,220	0,000	1,000
	NMM	0,000	1,000	0,562	0,197
A05	Data asli	1,500	6,000	4,311	1,342
	NZS	-2,095	1,258	0,000	1,000
	NMM	0,000	1,000	0,625	0,298
A06	Data asli	0,000	130,000	43,186	29,044
	NZS	-1,487	2,989	0,000	1,000
	NMM	0,000	1,000	0,332	0,223

Dataset asli memiliki variasi yang signifikan dalam nilai minimum, maksimum, rata-rata, dan deviasi standar di antara atribut-atributnya. Setelah proses normalisasi, nilai-nilai ini menjadi lebih seragam. Dalam metode Z-Score, nilai rata-rata (μ) dari atribut menjadi 0, sementara deviasi standarnya (σ) menjadi 1. Sedangkan pada metode Min-Max, nilai minimum dari semua atribut diubah menjadi 0, dan nilai maksimumnya menjadi 1.

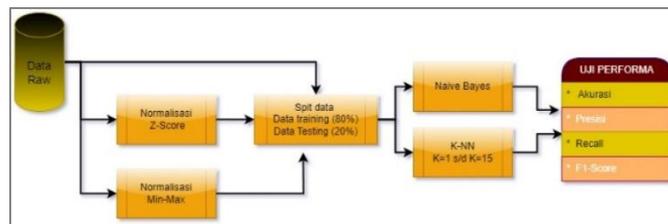
Ketiga dataset, termasuk dataset asli dan dataset hasil normalisasi menggunakan kedua teknik yang disebutkan, digunakan dalam proses pemodelan untuk memperoleh wawasan mengenai perbedaan kinerja yang akan dievaluasi pada tahap evaluasi

3.3 Modeling

Tahap pemodelan mencakup pemanfaatan teknik klasifikasi data mining, khususnya algoritma naïve Bayes dan K-NN, untuk membangun model prediktif dalam menentukan lokasi yang paling sesuai untuk layanan kependudukan dan pencatatan sipil *mobile*. Langkah penting dalam memastikan kemampuan model terletak pada pemilihan algoritma yang tepat. Alat bantu berupa aplikasi RapidMiner digunakan untuk memfasilitasi prosedur ini.

Pada tahap ini menginduksi data mining model (DMM) di RapidMiner menggunakan data yang telah disiapkan pada tahapan *Data preparation*, dengan menggunakan dua algoritma klasifikasi yakni Naïve bayes dan K-NN menggunakan ketiga jenis dataset kelayakan yakni tanpa normalisasi (data asli), NZS, dan NMM. Setiap dataset dibagi menjadi dua bagian dengan komposisi 80% (140 record) data latih dan 20% (35 record) data uji.

Gambaran umum pemodelan seperti terlihat pada gambar 2.



Gambar 2. Gambaran umum pemodelan

1) Naïve bayes

Naïve bayes. sesuai dengan prinsipnya, metode klasifikasi NB dikenal dengan asumsi "independensi bersyarat" antara variabel-variabel yang diberikan kelas[37], NB melakukan pengklasifikasian data dengan mempertimbangkan probabilitas dan statistik [38], [39] dari setiap nilai atribut terhadap nilai kelas target yang ingin dicapai. Dengan menggunakan theorema bayes[12] seperti persamaan (1):

$$P(c|x) = \frac{P(x|c)*P(c)}{P(x)} \tag{1}$$

Dimana $P(c|x)$ mewakili *probabilitas posterior*, $P(x|c)$ menunjukkan *probabilitas likelihood*, $P(c)$ adalah *probabilitas prior*, dan $P(x)$ adalah *probabilitas evidence*

Dalam dataset yang memiliki prediktor yang memiliki nilai kontinu, rumus yang digunakan untuk menentukan kemungkinan $P(x|c)$ adalah seperti dalam persamaan (2)

$$P(x_i|y_k) = \frac{1}{\sqrt{\sigma_k 2\pi}} \exp\left(-\frac{(x_i - \mu_k)^2}{2\sigma_k}\right) \tag{2}$$

Dimana $P(x_i|y_k)$ adalah *probabilitas likelihood*, $1/\sqrt{(\sigma_k 2\pi)}$ adalah faktor normalisasi yang digunakan untuk memastikan bahwa probabilitas yang dihitung berada dalam rentang antara 0 dan 1, dan $\exp(-((x_i - \mu_k)/(2\sigma_k))^2)$ adalah fungsi eksponensial yang mengukur seberapa jauh x_i dari μ_k dalam satuan deviasi baku (σ_k).

Setelah dilakukan pemodelan dengan menggunakan RapidMiner pada dataset asli, Dataset NZS dan NMM. Didapatkan jumlah prediksi yang sesuai dengan dunia nyata adalah 33 dari 35 data uji. Data hasil pemodelan naïve bayes menggunakan Rapidminer sebagaimana pada Tabel 7

Tabel 7. Data confusion matrix klasifikasi menggunakan Naïve Bayes

		Aktual								
		Dataset Asli			Dataset NZS			Dataset NMM		
		T-L	L	S-L	T-L	L	S-L	T-L	L	S-L
Prediksi	T-L	9	1	0	9	1	0	9	1	0
	L	1	16	0	1	16	0	1	16	0
	S-L	0	0	8	0	0	8	0	0	8

Tabel 7 dapat dilihat bahwa proses normlisasi tidak mempengaruhi hasil klasifikasi dari Naïve bayes, hal ini sebenarnya wajar karena dalam proses naïve bayes khususnya gaussian NB terdapat proses normalisasi, sehingga Normalisasi tidak diperlukan dalam algoritma NB. Dari hasil pemodelan dapat disimpulkan bahwa naïve bayes dapat memprediksi dengan benar 9 data kelas TIDAK LAYAK, 16 data kelas LAYAK dan 8 data kelas SANGAT LAYAK, Sementara 2 data diprediksi yang salah yaitu. 1 data pada kelas TIDAK LAYAK diprediksi LAYAK, dan 1 data kelas LAYAK diprediksi sebagai TIDAK LAYAK

2) K-NN

K-Nearest Neighbor (K-NN) adalah metode klasifikasi yang bekerja dengan mengukur kesamaan antara data baru dan data yang ada (data latih) [15] dan mengambil mayoritas kelas dari sejumlah k tetangga terdekat sebagai kelas target) [40]–[42]. Banyak metode perhitungan jarak yang digunakan untuk K-NN, seperti *Euclidean Distance*, *Manhattan Distance*, *Chebyshev Distance*, *Minkowsky Distance* [43]. Pada penelitian ini digunakan metode perhitungan jarak *Euclidean Distance* dengan rumus:

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

Dimana $d(x, y)$ adalah jarak *Euclidean* antara dua vektor x (data baru) dan y (data latih) dan x_i dan y_i mewakili atribut.

Pemodelan algoritma K-NN dengan variasi nilai dari k=1 hingga k=15 menggunakan alat bantu aplikasi RapidMiner, dan hasilnya direkap, sebagaimana terlihat Tabel 8.

Hasil pemodelan menunjukkan bahwa nilai prediksi yang tepat tertinggi terdapat pada k=1 dan k=2 ketika menggunakan dataset yang telah dinormalisasi menggunakan metode Normalisasi Min-Max (NMM), dengan jumlah prediksi yang tepat mencapai 34 dari total 35 data (97.14%). Sebaliknya, nilai prediksi terendah terlihat pada k=11 dan k=13 ketika menggunakan dataset asli, di mana hanya 27 dari 35 data yang diprediksi dengan benar (77,14%).

Tabel 8. Data Confusion matrix klasifikasi menggunakan K-NN

			Aktual								
			Dataset Asli			Dataset NZS			Dataset NMM		
			T-L	L	S-L	T-L	L	S-L	T-L	L	S-L
Prediksi	K=1	T-L	9	1	0	9	1	0	9	0	0
		L	1	16	0	1	16	0	1	17	0
		S-L	0	0	8	0	0	8	0	0	8
	K=2	T-L	9	1	0	9	1	0	9	0	0
		L	1	16	0	1	16	0	1	17	0
		S-L	0	0	8	0	0	8	0	0	8
	K=3	T-L	9	1	0	8	0	0	8	0	0
		L	1	14	0	2	17	0	2	17	1
		S-L	0	2	8	0	0	8	0	0	7
	K=4	T-L	9	1	0	8	0	0	8	0	0
		L	1	14	0	2	17	0	2	17	0
		S-L	0	2	8	0	0	8	0	0	8
	K=5	T-L	7	1	0	8	1	0	8	0	0
		L	3	14	0	2	16	0	2	17	0
		S-L	0	2	8	0	0	8	0	0	8
K=...	
K=15	T-L	7	1	0	8	1	0	8	1	0	
	L	3	14	1	2	16	0	2	16	0	
	S-L	0	2	7	0	0	8	0	0	8	

3.4 Evaluation

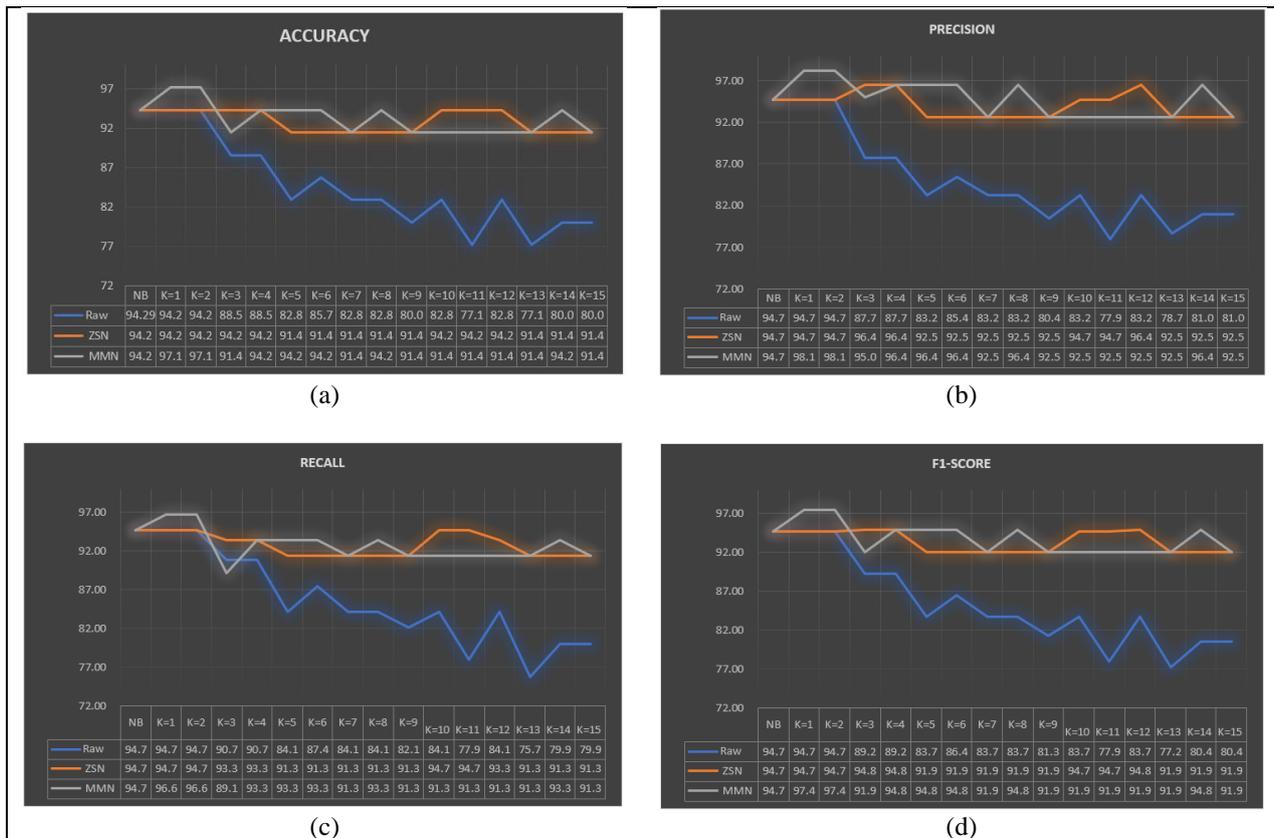
Proses ini mencakup pengujian performa model yang telah dikembangkan untuk memastikan tingkat akurasi yang optimal. Hasil dari evaluasi akan digunakan untuk melakukan perbandingan performa antara algoritma Naïve Bayes dan K-NN, yang akan membantu dalam menentukan algoritma yang paling sesuai untuk menentukan lokasi pelayanan kependudukan *mobile*. Evaluasi ini menggunakan tabel *confusion matrix*.

Tabel 9. Confusion matrix multi-class

		Aktual			Class Precision
		Kelas	A	B	
Prediksi	A	TP _A	E _{BA}	E _{CA}	
	B	E _{AB}	TP _B	E _{CB}	
	C	E _{AC}	E _{BC}	TP _C	
Class Recall					

Confusion matrix[44] (Tabel 9) terdiri dari baris dan kolom yang menggambar hasil prediksi dari aktual yang menyatakan *True Positive* (TP) yaitu jumlah data yang benar diprediksi sebagai positif (positif sesuai dengan kenyataan), *True Negative* (TN) yaitu Jumlah data yang benar diprediksi sebagai negatif (negatif sesuai dengan kenyataan) Contoh TNA=TPB+ECB+EBC+TPC, TNB=TPA+ECA+EAC+TPC, dan TNC=TPA+EBA+EAB+TPB, *False Positive* (FP) yaitu Jumlah data yang salah diprediksi sebagai positif, Contoh: FP_A=E_{BA}+E_{CA}, FP_B=E_{AB}+E_{CB}, dan FP_C= E_{AC}+E_{BC}, dan *False Negative* (FN): Jumlah data yang salah diprediksi sebagai negatif, Contoh FN_A = E_{AB}+E_{AC}, FN_B=E_{BA}+E_{BC}, dan FN_C=E_{CA}+E_{CB}

Dari analisis confusion matrix pemodelan menggunakan algoritma Naïve bayes Tabel 7 dan K-NN pada Tabel 8 dapat digambarkan tingkat *accuracy*, *precision*, *recall* dan *F1-Score* [45] seperti pada gambar 3 (a-d):



Gambar 3 Grafik (a) Akurasi, (b) Presisi, (c) Recall dan (d) F1-Score dari algoritma Naïve Bayes dan K-NN (k=1 sampai dengan k=15)

1) *Accuracy*.

Dari analisis confusion matrix diketahui akurasi tertinggi 97.14% dan tercapai pada dua kondisi: yakni menggunakan algoritma K-NN dengan nilai K=1 dan K=2 digunakan pada dataset NMM. Sedangkan akurasi terendah adalah 77.14% dan terjadi ketika K-NN dengan nilai K=11 dan K=13 digunakan pada Dataset Asli (tanpa normalisasi). Model Naive Bayes memiliki akurasi yang stabil sebesar 94.29% pada semua jenis dataset.

2) *Precision*

Nilai Presisi Tertinggi adalah 98.15% dan tercapai ketika menggunakan algoritma K-NN dengan nilai K=1 dan K=2 pada dataset NMM. Sedangkan Presisi terendah adalah 77.99% dan terjadi ketika menggunakan algoritma K-NN dengan nilai K=11 digunakan pada Dataset Asli (tanpa normalisasi). Algoritma Naive Bayes memiliki presisi yang sama sebesar 94.71% pada semua jenis dataset.

3) *Recall*

Nilai *Recall* (Sensitivitas) Tertinggi adalah 96.67% dan tercapai ketika menggunakan algoritma K-NN dengan nilai K=1 dan K=2 pada dataset NMM. Sedangkan *Recall* (Sensitivitas) Terendah adalah 75.78% dan terjadi ketika menggunakan algoritma K-NN dengan nilai K=13 digunakan pada Dataset Asli (tanpa normalisasi). Algoritma Naive Bayes memiliki recall yang stabil sebesar 94.71% pada semua jenis dataset.

4) *F1-Score*

Nilai F1-Score tertinggi adalah 97.40% dan tercapai ketika menggunakan algoritma K-NN dengan nilai K=1 dan K=2 pada dataset NMM. Sedangkan F1-Score Terendah adalah 77.23% dan terjadi ketika menggunakan algoritma K-NN dengan nilai K=13 pada Dataset Asli (tanpa normalisasi). Algoritma Naive Bayes memiliki F1-Score yang stabil sebesar 94.71% pada semua jenis dataset.

4. KESIMPULAN

Layanan kependudukan *mobile* bisa menjadi solusi efektif dan efisien untuk layanan masyarakat khususnya dalam pelayanan kependudukan dan pencatatan sipil, namun pemilihan lokasi pelayanan harus dilakukan secara hati-hati mengingat mencakup banyak desa yang perlu dijangkau. Penentuan lokasi desa selama ini seringkali dilakukan secara manual, menyebabkan 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 94,71%. Performa K-NN berfluktuasi tergantung pada nilai K dan jenis dataset yang digunakan. K-NN mencapai tingkat akurasi tertinggi sebesar 97,14% saat menggunakan dataset NMM pada K=1 dan K=2, dengan rata-rata presisi 98,1% dan nilai recall 96,6%. Berdasarkan temuan ini, peneliti merekomendasikan penggunaan algoritma K-NN dalam menentukan lokasi layanan kependudukan *mobile*, dengan persyaratan untuk melakukan normalisasi data menggunakan NMM terlebih dahulu, dan nilai K yang optimal adalah 1 atau 2.

Ucapan Terimakasih

Terima kasih kepada Program Studi Magister Informatika dan Universitas Ahmad Dahlan yang telah memfasilitasi dan mendanai penelitian ini.

DAFTAR PUSTAKA

- [1] R. Indonesia, "Undang Undang Republik Indonesia No. 23 Tahun 2006 Tentang Administrasi Kependudukan." Jakarta, 2006.
- [2] Kemendagri, "Peraturan Menteri Dalam Negeri Republik Indonesia Nomor 7 Tahun 2019 tentang Pelayanan Administrasi Kependudukan Secara Daring." Jakarta, 2019.
- [3] Kemendagri, "Peraturan Menteri Dalam Negeri No. 109 Tahun 2019 Tentang Formulir Dan Buku Yang Digunakan Dalam Administrasi Kependudukan." Jakarta, 2019.
- [4] Kemendagri, "Peraturan Menteri Dalam Negeri Republik Indonesia Nomor 72 Tahun 2022 Tentang Standar dan Spesifikasi Perangkat Keras, Perangkat Lunak, Dan Blangko Kartu Tanda Penduduk Elektronik Serta Penyelenggaraan Identitas Kependudukan Digital." Jakarta, p. 32, 2022.
- [5] Dinas Komunikasi Dan Informatika Kabupaten Alor, "Geografis," *alorkab.go.id*, 2022. <http://www.alorkab.go.id/new/index.php/profil/geografis1> (accessed Aug. 27, 2023).
- [6] E. M. F. El Houbby, "A survey on applying machine learning techniques for management of diseases," *J. Appl. Biomed.*, vol. 16, no. 3, pp. 165–174, Aug. 2018, doi: 10.1016/j.jab.2018.01.002.
- [7] S. Mishra, R. Kumar, S. K. Tiwari, and P. Ranjan, "Machine learning approaches in the diagnosis of infectious diseases: a review," *Bull. Electr. Eng. Informatics*, vol. 11, no. 6, pp. 3509–3520, 2022, doi: 10.11591/eei.v11i6.4225.
- [8] M. E. Febrian, F. X. Ferdinan, G. P. Sendani, K. M. Suryanigrum, and R. Yunanda, "Diabetes prediction using supervised machine learning," *Procedia Comput. Sci.*, vol. 216, no. 2022, pp. 21–30, 2023, doi: 10.1016/j.procs.2022.12.107.
- [9] M. M. Hossin, F. M. Javed Mehedi Shamrat, M. R. Bhuiyan, R. Akter Hira, T. Khan, and S. Molla, "Breast cancer detection: an effective comparison of different machine learning algorithms on the Wisconsin dataset," *Bull. Electr. Eng. Informatics*, vol. 12, no. 4, pp. 2446–2456, 2023, doi: 10.11591/eei.v12i4.4448.
- [10] K. M. Almustafa, "Prediction of chronic kidney disease using different classification algorithms," *Informatics Med. Unlocked*, vol. 24, p. 100631, 2021, doi: 10.1016/j.imu.2021.100631.
- [11] R. Safdari, A. Deghatipour, M. Gholamzadeh, and K. Maghooli, "Applying data mining techniques to classify patients with suspected hepatitis C virus infection," *Intell. Med.*, no. December 2021, 2022, doi: 10.1016/j.imes.2021.12.003.
- [12] C. Patgiri and A. Ganguly, "Adaptive thresholding technique based classification of red blood cell and sickle cell using Naïve Bayes Classifier and K-nearest neighbor classifier," *Biomed. Signal Process. Control*, vol. 68, no. July 2020, p. 102745, 2021, doi: 10.1016/j.bspc.2021.102745.
- [13] D. D. Khudhur and S. D. Khudhur, "The classification of autism spectrum disorder by machine learning methods on multiple datasets for four age groups," *Meas. Sensors*, vol. 27, no. March, p. 100774, 2023, doi: 10.1016/j.measen.2023.100774.
- [14] K. Polat, S. Şahan, and S. Güneş, "Automatic detection of heart disease using an artificial immune recognition system (AIRS) with fuzzy resource allocation mechanism and k-nn (nearest neighbour) based weighting preprocessing," *Expert Syst. Appl.*, vol. 32, no. 2, pp. 625–631, 2007, doi: 10.1016/j.eswa.2006.01.027.
- [15] A. Yudhana, A. Muslim, D. E. Wati, I. Puspitasari, A. Azhari, and M. M. Mardhia, "Human emotion recognition based on EEG signal using fast fourier transform and K-Nearest neighbor," *Adv. Sci. Technol. Eng. Syst.*, vol. 5, no. 6, pp. 1082–1088, 2020, doi: 10.25046/aj0506131.
- [16] S. Lahmiri, C. Tadj, C. Gargour, and S. Bekiros, "Optimal tuning of support vector machines and k-NN algorithm by using Bayesian optimization for newborn cry signal diagnosis based on audio signal processing features," *Chaos, Solitons and Fractals*, vol. 167, no. December 2022, p. 112972, 2023, doi: 10.1016/j.chaos.2022.112972.
- [17] I. Konovalenko and A. Ludwig, "Generating decision support for alarm processing in cold supply chains using a hybrid k-NN algorithm," *Expert Syst. Appl.*, vol. 190, no. December 2020, 2022, doi: 10.1016/j.eswa.2021.116208.
- [18] M. P. Vaishnav, K. Suganya Devi, P. Srinivasan, and G. Arutperumjothi, "Detection and classification of groundnut leaf diseases using KNN classifier," *2019 IEEE Int. Conf. Syst. Comput. Autom. Networking, ICSCAN 2019*, pp. 1–5, 2019, doi: 10.1109/ICSCAN.2019.8878733.
- [19] Marisa F et al., "Pengukuran Tingkat Kematangan Kopi Arabika Menggunakan Algoritma K-Nearest Neighbor," *JIMPJurnal Inform. Merdeka Pasuruan*, vol. 6, no. 3, pp. 1–5, 2021. [Online]. Available: <http://ejournal.unmerpas.ac.id/index.php/informatika/article/view/280/122>
- [20] A. Yudhana, D. Sulisty, and I. Mufandi, "GIS-based and Naïve Bayes for nitrogen soil mapping in Lendah, Indonesia," *Sens. Bio-Sensing Res.*, vol. 33, pp. 100435:1-100435:11, 2021, doi: 10.1016/j.sbsr.2021.100435.
- [21] A. Yudhana, A. D. Cahyo, L. Y. Sabila, A. C. Subrata, and I. Mufandi, "Spatial distribution of soil nutrient content for sustainable rice agriculture using geographic information system and Naïve Bayes classifier," *Int. J. Smart Sens. Intell. Syst.*, vol. 16, no. 1, pp. 1–14, 2023, doi: 10.2478/ijssis-2023-0001.
- [22] T. Setiadi, F. Noviyanto, H. Hardianto, A. Tarmuji, A. Fadlil, and M. Wibowo, "Implementation of naïve bayes method in food crops planting recommendation," *Int. J. Sci. Technol. Res.*, vol. 9, no. 2, pp. 4750–4755, 2020.
- [23] A. Yudhana, R. Umar, and S. Saputra, "Fish Freshness Identification Using Machine Learning: Performance Comparison of k-NN and Naïve Bayes Classifier," *J. Comput. Sci. Eng.*, vol. 16, no. 3, pp. 153–164, 2022, doi: 10.5626/JCSE.2022.16.3.153.
- [24] K. Mitrovic, D. Milosevic, and M. Greconici, "Comparison of Machine Learning Algorithms for Shelter Animal Classification," *SACI 2019 - IEEE 13th Int. Symp. Appl. Comput. Intell. Informatics, Proc.*, pp. 211–216, 2019, doi: 10.1109/SACI46893.2019.9111575.
- [25] A. Yudhana, I. Riadi, and M. R. Djou, "Pengembangan Layanan Kependudukan Dan Pencatatan Sipil Menggunakan Algoritma Naïve Bayes," *JURIKOM (Jurnal Ris. Komputer)*, vol. 9, no. 4, p. 1062, 2022, doi: 10.30865/jurikom.v9i4.4515.
- [26] A. Yudhana, I. Riadi, and M. R. Djou, "Determining eligible villages for mobile services using k- NN algorithm," vol. 15, no. 1, pp. 11–20, 2023, doi: <http://dx.doi.org/10.30865/jurikom.v9i4.4515>.
- [27] N. Dengen, "Comparison Performance of C4.5, Naïve Bayes And K-Nearest Neighbor in Determination Drug Rehabilitation," *Int. Conf. Sci. Inf. Technol.*, pp. 112–117, 2019, doi: DOI:10.1109/ICSITech46713.2019.8987455.
- [28] S. Yonesa, S. M. Nasution, and R. A. Nugrahaeni, "Path Determination with Classification from Data Mining Using C4.5 Algorithm," *2018 Int. Conf. Inf. Technol. Syst. Innov. ICITSI 2018 - Proc.*, pp. 22–25, 2018, doi: 10.1109/ICITSI.2018.8696020.
- [29] S. Sunardi, A. Yudhana, and S. A. Wijaya, "Face Detection Analysis of Digital Photos Using Mean Filtering Method," *Int. J. Artif. Intell. Res.*, vol. 6, no. 2, 2022, doi: 10.29099/ijair.v6i2.307.
- [30] C. Schröer, F. Kruse, and J. M. Gómez, "A systematic literature review on applying CRISP-DM process model," *Procedia Comput. Sci.*, vol. 181, no.



- 2019, pp. 526–534, 2021, doi: 10.1016/j.procs.2021.01.199.
- [31] S. Huber, H. Wiemer, D. Schneider, and S. Ihlenfeldt, “DMME: Data mining methodology for engineering applications - A holistic extension to the CRISP-DM model,” *Procedia CIRP*, vol. 79, pp. 403–408, 2019, doi: 10.1016/j.procir.2019.02.106.
- [32] J. A. Solano, D. J. Lancheros Cuesta, S. F. Umaña Ibáñez, and J. R. Coronado-Hernández, “Predictive models assessment based on CRISP-DM methodology for students performance in Colombia - Saber 11 Test,” *Procedia Comput. Sci.*, vol. 198, no. 2020, pp. 512–517, 2021, doi: 10.1016/j.procs.2021.12.278.
- [33] A. Morais, H. Peixoto, C. Coimbra, A. Abelha, and J. Machado, “Predicting the need of Neonatal Resuscitation using Data Mining,” *Procedia Comput. Sci.*, vol. 113, pp. 571–576, 2017, doi: 10.1016/j.procs.2017.08.287.
- [34] A. Abubakar and A. B. Garko, “A Predictive Model for Network Intrusion Detection System Using Deep Neural Network,” *Dutse J. Pure Appl. Sci.*, vol. 7, no. 3a, pp. 113–128, 2022, doi: 10.4314/dujopas.v7i3a.12.
- [35] Y. A. Setianto, K. Kusriani, and H. Henderi, “Penerapan Algoritma K-Nearest Neighbour Dalam Menentukan Pembinaan Koperasi Kabupaten Kotawaringin Timur,” *Creat. Inf. Technol. J.*, vol. 5, no. 3, pp. 232–241, 2019, doi: 10.24076/citec.2018v5i3.179.
- [36] D. Singh and B. Singh, “Investigating the impact of data normalization on classification performance,” *Appl. Soft Comput.*, vol. 97, no. B, p. 105524, 2020, doi: 10.1016/j.asoc.2019.105524.
- [37] A. Pérez, P. Larrañaga, and I. Inza, “Supervised classification with conditional Gaussian networks: Increasing the structure complexity from naive Bayes,” *Int. J. Approx. Reason.*, vol. 43, no. 1, pp. 1–25, 2006, doi: 10.1016/j.ijar.2006.01.002.
- [38] Rovidatul, Y. Yunus, and G. W. Nurcahyo, “Perbandingan algoritma c4.5 dan naive bayes dalam prediksi kelulusan mahasiswa,” *J. CoSciTech (Computer Sci. Inf. Technol.)*, vol. 4, no. 1, pp. 193–199, 2023, doi: 10.37859/coscitech.v4i1.4755.
- [39] Fitri Wulandari, Elin Haerani, Muhammad Fikry, and Elvia Budianita, “Analisis sentimen larangan penggunaan obat sirup menggunakan algoritma naive bayes classifier,” *J. CoSciTech (Computer Sci. Inf. Technol.)*, vol. 4, no. 1, pp. 88–96, 2023, doi: 10.37859/coscitech.v4i1.4781.
- [40] V. Tsenev and M. Ivanova, “Statistical and machine learning approach for evaluation of control systems for automatic production lines,” *Bull. Electr. Eng. Informatics*, vol. 11, no. 5, pp. 2527–2536, 2022, doi: 10.11591/eei.v11i5.3664.
- [41] R. Ritzkal *et al.*, “K-nearest neighbor algorithm analysis for path determination in network simulation using software defined network,” *Bull. Electr. Eng. Informatics*, vol. 12, no. 4, pp. 2388–2400, 2023, doi: 10.11591/eei.v12i4.4868.
- [42] R. Gunawan, R. Septiadi, F. Apri Wenando, H. Mukhtar, and Syahril, “K-Nearest Neighbor (KNN) untuk Menganalisis Sentimen terhadap Kebijakan Merdeka Belajar Kampus Merdeka pada Komentar Twitter,” *J. CoSciTech (Computer Sci. Inf. Technol.)*, vol. 3, no. 2, pp. 152–158, 2022, doi: 10.37859/coscitech.v3i2.3841.
- [43] A. N. Sadrabadi, S. M. Znjirchi, H. Z. A. Abadi, and A. Hajimoradi, “An optimized K-Nearest Neighbor algorithm based on Dynamic Distance approach,” *6th Iran. Conf. Signal Process. Intell. Syst. ICSPIS 2020*, 2020, doi: 10.1109/ICSPIS51611.2020.9349582.
- [44] A. Tharwat, “Classification assessment methods,” *Appl. Comput. Informatics*, vol. 17, no. 1, pp. 168–192, 2018, doi: 10.1016/j.aci.2018.08.003.
- [45] A. Sirisha, K. Chaitanya, K. V. S. S. R. Krishna, and S. S. Kanumalli, “Intrusion detection models using supervised and unsupervised algorithms - A comparative estimation,” *Int. J. Saf. Secur. Eng.*, vol. 11, no. 1, pp. 51–58, 2021, doi: 10.18280/ijss.110106.