



Prediction of Birth Rates in Different Age Groups of Mothers Using the K-Nearest Neighbor (K-NN) Method

Prediksi Angka Kelahiran dalam Berbagai Kelompok Umur Ibu Menggunakan Metode K-Nearest Neighbor

Vinka Syafana^{1*}, Shofa Shofiah Hilabi², Elfina Novalia³, Baenil Huda⁴

^{1,2,3,4}Program Studi Sistem Informasi, Fakultas Ilmu Komputer,
Universitas Buana Perjuangan Karawang, Indonesia

E-Mail: ¹si20.vinkasyafana@mhs.ubpkarawang.ac.id, ²shofa.hilabi@ubpkarawang.ac.id,
³elfinanovalia@ubpkarawang.ac.id, ⁴baenil88@ubpkarawang.ac.id

Received May 7th 2024; Revised Jun 27th 2024; Accepted Jul 10th 2024
Corresponding Author: Vinka Syafana

Abstract

This study aims to develop a prediction of birth rates in different age groups of mothers through the K-Nearest Neighbor (K-NN) method approach in an information system. The research focused on analysing birth dynamics in the context of maternal age groups, which is critical information for health policy planning and social development. The K-NN method is used as the main analytical approach for forecasting birth rates, utilising patterns of similarity in maternal age group characteristics. The integration of the K-NN method in the information system enables more efficient data management and analysis to support family planning policies. The data used includes demographic, economic, and social variables that may affect birth rates. The results of birth rate prediction are expected to provide deep insights into dynamic changes in the birth structure, enabling a better understanding to direct health policies and more targeted intervention strategies. This research presents a contribution to the development of information systems that can support predictive analyses in the context of births. The implications of the findings are relevant for health policy, demography and family planning, and can help design more effective and targeted intervention measures.

Keyword: Birth Rate, Information System, K-Nearest Neighbor, Maternal Age Group

Abstrak

Studi ini bertujuan untuk mengembangkan prediksi angka kelahiran dalam berbagai kelompok umur ibu melalui pendekatan metode K-Nearest Neighbors (K-NN) dalam suatu sistem informasi. Penelitian ini difokuskan pada analisis dinamika kelahiran dalam konteks kelompok umur ibu, yang merupakan informasi kritis dalam perencanaan kebijakan kesehatan dan pengembangan sosial. Metode K-NN digunakan sebagai pendekatan analisis utama untuk meramalkan angka kelahiran, memanfaatkan pola kemiripan dalam karakteristik kelompok umur ibu. Integrasi metode K-NN dalam sistem informasi memungkinkan pengelolaan dan analisis data yang lebih efisien untuk mendukung kebijakan perencanaan keluarga. Data yang digunakan mencakup variabel demografis, ekonomi, dan sosial yang dapat memengaruhi tingkat kelahiran. Hasil prediksi angka kelahiran diharapkan dapat memberikan wawasan mendalam tentang perubahan dinamika dalam struktur kelahiran, memungkinkan pemahaman yang lebih baik untuk mengarahkan kebijakan kesehatan dan strategi intervensi yang lebih terarah. Penelitian ini menghadirkan kontribusi pada pengembangan sistem informasi yang dapat mendukung analisis prediktif dalam konteks kelahiran. Implikasi temuan ini relevan untuk kebijakan kesehatan, demografi, dan perencanaan keluarga, serta dapat membantu merancang langkah-langkah intervensi yang lebih efektif dan tepat sasaran.

Kata Kunci: Angka Kelahiran, Kelompok Umur Ibu, K-Nearest Neighbor, Sistem Informasi

1. PENDAHULUAN

Peningkatan kualitas sistem informasi menjadi krusial dalam era modern ini, terutama ketika berbicara tentang perencanaan kebijakan kesehatan dan pengembangan sosial. Salah satu aspek penting dalam konteks ini adalah prediksi angka kelahiran, yang menjadi indikator utama dalam perencanaan keluarga dan pembangunan masyarakat. Sebagai tanggapan terhadap kompleksitas dan dinamika perubahan sosial, teknologi

informasi dan metode analisis data semakin menjadi peran penting dalam menggali wawasan mendalam mengenai kelahiran dalam berbagai kelompok umur ibu [1].

Studi ini bertujuan untuk mengintegrasikan metode K- Nearest Neighbors (K-NN) dalam pengembangan sistem informasi untuk memprediksi angka kelahiran dalam kelompok umur ibu. Pendekatan ini diadopsi karena kemampuannya dalam mengidentifikasi pola dan karakteristik yang mungkin sulit terlihat dengan metode analisis tradisional [2]. K-NN menawarkan solusi yang efisien untuk memproses data yang kompleks, memungkinkan pengambilan keputusan yang lebih baik dalam perencanaan keluarga dan kebijakan kesehatan [3].

Periode studi ini difokuskan pada tahun 2020, mengakui bahwa dinamika sosial dan ekonomi pada periode tersebut dapat memberikan dampak signifikan pada keputusan untuk memiliki anak. Penggunaan metode K-NN dalam sistem informasi dapat memberikan keunggulan analisis prediktif dalam meramalkan angka kelahiran, mempertimbangkan variabel demografis, ekonomi, dan sosial yang beragam [4]. Kelompok umur ibu dianggap sebagai elemen kritis dalam analisis ini, mengingat perbedaan karakteristik dan tantangan yang dihadapi oleh setiap kelompok umur [5]. Dengan merinci analisis pada tingkat kelompok umur, diharapkan hasil prediksi angka kelahiran dapat memberikan pemahaman yang lebih kaya dan kontekstual tentang tren kelahiran dalam masyarakat.

Pentingnya integrasi metode K-NN dalam sistem informasi tidak hanya terletak pada efektivitas analisis data, tetapi juga pada kemampuannya untuk memberikan wawasan yang lebih mendalam tentang perubahan dinamis dalam struktur kelahiran [6]. Analisis prediktif ini bukan hanya memberikan gambaran statistik, tetapi juga mempertimbangkan faktor-faktor kontekstual yang dapat memengaruhi keputusan keluarga [7]. Penelitian ini tidak hanya mengkaji literatur terkait algoritma K-Nearest Neighbor, tetapi juga meninjau penelitian terdahulu yang membahas integrasi metode K-NN dalam sistem informasi untuk analisis data demografi dan perencanaan keluarga, hal ini dilakukan untuk memperkuat landasan teoretis dan memberikan gambaran komprehensif tentang penelitian yang dilakukan sebelumnya. Beberapa penelitian terdahulu telah menunjukkan efektivitas metode K-NN dalam memprediksi kelahiran [8, 9, 10].

Penelitian oleh Lit et al. (2020) menggunakan metode K-NN untuk memprediksi kelahiran prematur dengan akurasi mencapai 92%. Penelitian lain oleh Chen et al. (2021) menggunakan metode K-NN untuk memprediksi tingkat kematian bayi dengan akurasi mencapai 88%, Rahman, M.S., & Chi, T. H (2022) peneliti ini menggunakan metode K-NN untuk memprediksi tingkat kematian neonatus di Bangladesh dengan akurasi mencapai 86%, penelitian ini menggunakan data dari 50.000 kelahiran neonatus dan menunjukkan bahwa metode K-NN dapat digunakan untuk mengidentifikasi faktor-faktor risiko kematian neonatus dengan tepat Singh & Kaur (2024) peneliti ini menganalisis data dari 300.000 kelahiran bayi di India untuk memprediksi tingkat kematian perinatal Algoritma K-NN digunakan untuk mengidentifikasi faktor-faktor risiko kematian perinatal, seperti usia ibu, riwayat kehamilan sebelumnya, dan komplikasi kehamilan. Hasil penelitian menunjukkan bahwa metode K-NN dapat mencapai akurasi prediksi sebesar 80%.

Dengan menyelidiki dan merancang sistem informasi yang mampu memadukan metode K-NN, studi ini berupaya memberikan kontribusi pada literatur penelitian di bidang sistem informasi, demografi, dan perencanaan keluarga [8]. Temuan dari penelitian ini diharapkan dapat memberikan pandangan baru dan relevan untuk mendukung perencanaan kebijakan yang lebih cerdas dan mendalam di bidang kesehatan dan pembangunan sosial [9]. Penggunaan algoritma K-NN yang dikombinasikan dengan teknik analisis data lainnya untuk menghasilkan prediksi yang lebih akurat dan terperinci tentang struktur kelahiran. Pertimbangan faktor-faktor kontekstual yang memengaruhi keputusan keluarga dalam perencanaan keluarga, seperti kondisi ekonomi, sosial, dan budaya. Pengembangan sistem informasi yang terintegrasi dan mudah digunakan oleh pemangku kepentingan di bidang kesehatan dan pembangunan sosial.

Adapun yang menjadi perbedaan penelitian terdahulu seperti peneliti yang dilakukan Lit et al, menggunakan metode K-NN untuk memprediksi kelahiran prematur, peneliti Rahman, M.S., & Chi, T.,H menggunakan metode K-NN untuk memprediksi kematian bayi, Peneliti Singh & Kaur menggunakan metode K-NN untuk memprediksi angka kelahiran bayi secara perenantal, sedangkan penelitian ini lebih berfokus kepada Prediksi Angka Kelahiran dalam Berbagai Kelompok Umur Ibu menggunakan metode K-NN.

2. METODE PENELITIAN

Berdasarkan temuan penelitian ini, pendekatan K-Nearest Neighbor (K-NN) akan dimasukkan ke dalam pengembangan sistem informasi yang akan meramalkan angka kelahiran pada kelompok umur ibu pada tahun 2020 [10]. Teknik stratified random digunakan untuk mengambil sampel dari populasi penelitian, yang mencakup kelompok usia ibu dari berbagai lokasi geografis. Hal ini dilakukan untuk menjamin bahwa pemilihan sampel mewakili kelompok usia. Banyak sumber berbeda, seperti lembaga kesehatan dan data survei demografi, akan dikonsultasikan untuk mengumpulkan data terkait, yang mencakup informasi demografi, ekonomi, dan sosial. Penggunaan metode K-NN untuk membangun model prediktif sekaligus mengoptimalkan parameter, seperti jumlah tetangga (K), melalui proses validasi silang akan dapat dilakukan [11]. Untuk memperoleh pemahaman yang lebih mendalam mengenai pola kelahiran di masyarakat, maka temuan prediksi tersebut akan dievaluasi pada tingkat kelompok umur ibu [12]. Untuk memberikan perspektif yang lebih

mendalam, aspek kontekstual seperti perubahan lingkungan sosial dan ekonomi akan dimasukkan. Validasi model akan dilakukan dengan bantuan data independen, dan hasilnya akan divalidasi melalui partisipasi ahli demografi dan kebijakan untuk menjamin bahwa temuan tersebut relevan dan diinterpretasikan dengan benar [13].

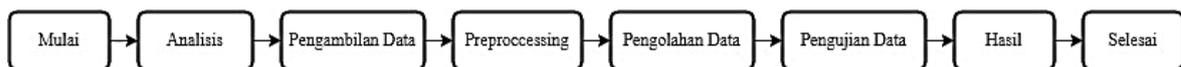
2.1. Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah data pengumpulan yang terkait dengan suatu topik atau domain tertentu. Data tersebut terdiri dari berbagai atribut atau variabel yang memiliki peran penting dalam analisis. Atribut atau variabel utama yang digunakan adalah metode K-NN atribut ini memiliki peran signifikan dalam penelitian dan memberikan informasi yang relevan terkait dengan topik yang sedang diteliti, selain atribut utama juga digunakan variabel kelas, variabel kelas ini memiliki peran sebagai target atau output yang ingin di prediksi atau diklasifikasikan.

Untuk analisis menggunakan metode K-Nearest Neighbor (K-NN), beberapa parameter-parameter yang digunakan adalah sebagai berikut:

1. Jumlah tetangga terdekat, parameter ini menentukan jumlah tetangga terdekat yang akan digunakan dalam proses klasifikasi atau prediksi.
2. Metrik jarak, parameter ini menentukan metode pengukuran jarak antara data yang akan di klasifikasikan dengan tetangga terdekat.
3. Bobot jarak, parameter ini menentukan bobot yang diberikan pada jarak antara data yang akan diklasifikasikan dengan tetangga terdekatnya

Dalam penelitian ini parameter-parameter K-NN yang digunakan adalah *Precision*, *Recall*, *Accuracy* nilai-nilai parameter ini dipilih berdasarkan pertimbangan kebutuhan penelitian.



Gambar 1. Alur Penelitian

Dalam penelitian ini data yang digunakan merupakan data Prediksi Angka Kelahiran dalam Berbagai Kelompok Umur Ibu dengan Menggunakan Metode K-NN. Selanjutnya data prediksi angka kelahiran tersebut dilakukan pra-proses untuk dapat menghasilkan data yang khusus untuk mtuk menjadi informasi yang baru dan bermanfaat [14]. Berikut merupakan penjelasan pengambilan data:

1. Tahap Pengambilan data: Data yang didapatkan dilakukan proses terlebih dahulu. Dengan adanya data selection ini proses dalam pengolahan data akan lebih baik dengansesuai tujuan penelitian yang akan dicapainya. Dalam hal ini peneliti menggunakan data selama satu tahun di tahun 2020.
2. Tahap *Pre-Processing*: sebelum proses data mining dapat dilaksanakan, perlu dilakukannya proses *Cleaning*. Data *Cleaning* mencakup antara lain membuang duplikasi data memeriksa data yang inkonsisten dan memperbaiki kesalahan dalam data.
3. Tahap Pengolahan data: setelah dilakukan *cleaning* data, maka kemudian akan Proses untuk pengolahan data berdasarkan algoritma yang sesuai yaitu dengan menggunakan *data mining*. Algoritma yang dilakukan dalam penelitian ini menggunakan metode K-Nearest Neighbors.
4. Tahap Pengujian data: setelah data di olah menggunakan K- Nearest Neighbors (K-NN), Kemudian data diuji menggunakan Rapidminer sebagai tools untuk memperoleh hasil yang selaras dengan perhitungan secara manual.
5. Tahap Hasil: Setelah dilakukannya pengujian dengan menggunakan RapidMiner sebagai tools, maka akan memperoleh hasil. Dengan hasil ini merupakan akhir dari proses penelitian ini.

2.2. Algoritma K-Nearest Neighbor (K-NN)

Metode K-Nearest Neighbor (K-NN) umumnya diterapkan dalam tugas klasifikasi data. Teknik ini berfokus pada pengelompokan objek berdasarkan data pembelajaran yang memiliki jarak terdekat atau nilai selisih kecil dengan objek yang sedang diklasifikasikan. K-NN merupakan suatu pendekatan yang mengandalkan algoritma supervised, di mana hasil klasifikasi instance baru ditentukan oleh mayoritas kategori pada K-NN [15].

Tujuan utama dari algoritma K-NN adalah mengklasifikasikan objek baru dengan mempertimbangkan atribut dan sampel pelatihan yang telah ada [16]. Prinsip dasar dari algoritma ini melibatkan pencarian nilai K dari data pelatihan untuk menentukan tetangga terdekat berdasarkan ukuran jarak. Selanjutnya, mayoritas nilai K dari tetangga terdekat akan menjadi dasar untuk menentukan kategori dari sampel yang akan diklasifikasikan berikutnya [17].

Metode K-Nearest Neighbors (K-NN) seringkali digunakan untuk tugas klasifikasi dalam teknik data mining. Meskipun demikian, algoritma ini juga dapat diterapkan untuk estimasi dalam prediksi data, menunjukkan fleksibilitasnya dalam berbagai konteks analisis data [18].

Adapun langkah-langkah dalam metode *K-NearestNeighbor* adalah sebagai berikut:

1. Menentukan nilai K
2. Menghitung jarak *Euclidean*

Rumus *Euclidean*:

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

Keterangan:

- d(x,y) : jarak antara data x ke data y
- x_i : nilai x_i pada data training
- y_i : nilai y_i pada data testing
- N : variable data
- I : dimensi data

3. Mengurutkan berdasarkan nilai *Euclidean* distance
4. Menentukan *k record* klasifikasi yang paling terdekat
5. Target *output* merupakan kelas yang mayoritas.

3. HASIL DAN PEMBAHASAN

Data yang diuraikan dalam penelitian ini berkaitan dengan angka kelahiran dalam berbagai kelompok umur ibu selama tahun 2020. Tabel 1 adalah kumpulan data sepanjang tahun 2020 yang dapat diakses dalam platform Badan Pusat Statistik.

Tabel 1. Dataset Awal

No	Kabupaten	Umur 15-19	Umur 20-24	Umur 25-29	Umur 30-34	Umur 35-39	Umur 40-44	Umur 45-49
1	Lampung Barat	49,21	134,4	138,34	102,28	58,25	15,43	3,93
2	Tanggamus	21,91	102,61	136,45	115,61	73,81	21,78	0,94
3	Lampung Selatan	29,56	109,38	140,8	108,05	57,43	19,26	0,59
4	Lampung Timur	47,99	107,51	118,37	94,19	66,72	18,34	2,31
5	Lampung Tengah	30,8	101,1	126,8	104	61	16,6	2,8
6	Lampung Utara	27,14	107,79	135,31	103,69	71	20,62	2,46
7	Way Kanan	34,2	118,8	128,55	101,76	61,84	17,2	2,53
8	Tulangbawang	37,4	108,62	121,87	107,39	59,22	15,68	2,65
9	Pesawaran	21,59	102,38	133,45	115,36	65,31	19,61	1,03
10	Pringsewu	19,6	90,7	146,2	113,6	67,9	17,6	1,7
11	Mesuji	50,29	112	142,91	103,56	58,26	14,75	0,86
12	Tulang Bawang Barat	33,2	122,2	123,5	96,1	55,7	16,3	2
13	Pesisir Barat	38,6	120,3	151,2	114,8	69,8	19,2	1,3
14	Kota Bandar Lampung	9,75	69,55	149,02	121,29	69,8	17,35	1,38
15	Kota Metro	10,6	78,7	138,2	109,6	63,6	18,4	1,3
16	Lampung	34,5	108,3	135,7	99,3	59,5	15,9	3,3

3.1. Dataset Algoritma K-NN

Data pengelompokan angka kelahiran dalam berbagai umur ibu ini diolah dan juga dikelompokkan dengan menggunakan algoritma data mining K-Nearest Neighbor (K-NN) untuk memprediksi menentukan data tersebut masuk kedalam kategori rendah, sedang dan tinggi. Selanjutnya proses dilanjutkan dengan pembagian data menjadi dua segmen. Peneliti memakai 16 data yang terdapat dalam file dataset selama tahun 2020. Dari 16 data tersebut akan dibagi menggunakan aturan 80/20 atau prinsip Pareto yaitu 13 data training dan 3 data testing yang akan dijadikan sebagai proses perhitungan manual, ditunjukkan pada tabel 2 dan tabel 3.

Tabel 2. Data Training

No	Kabupaten	Umur 15-19	Umur 20-24	Umur 25-29	Umur 30-34	Umur 35-39	Umur 40-44	Umur 45-49	Kategori
1	Lampung Barat	49,21	134,4	138,34	102,28	58,25	15,43	3,93	Tinggi
2	Tanggamus	21,91	102,61	136,45	115,61	73,81	21,78	0,94	Sedang
3	Lampung Selatan	29,56	109,38	140,8	108,05	57,43	19,26	0,59	Rendah
4	Lampung Timur	47,99	107,51	118,37	94,19	66,72	18,34	2,31	Rendah
5	Lampung Tengah	30,8	101,1	126,8	104	61	16,6	2,8	Rendah

No	Kabupaten	Umur 15-19	Umur 20-24	Umur 25-29	Umur 30-34	Umur 35-39	Umur 40-44	Umur 45-49	Kategori
6	Lampung Utara	27,14	107,79	135,31	103,69	71	20,62	2,46	Sedang
7	Way Kanan	34,2	118,8	128,55	101,76	61,84	17,2	2,53	Rendah
8	Tulangbawang	37,4	108,62	121,87	107,39	59,22	15,68	2,65	Rendah
9	Pesawaran	21,59	102,38	133,45	115,36	65,31	19,61	1,03	Rendah
10	Pringsewu	19,6	90,7	146,2	113,6	67,9	17,6	1,7	Rendah
11	Mesuji	50,29	112	142,91	103,56	58,26	14,75	0,86	Sedang
12	Tulang Bawang Barat	33,2	122,2	123,5	96,1	55,7	16,3	2	Rendah
13	Pesisir Barat	38,6	120,3	151,2	114,8	69,8	19,2	1,3	Tinggi

Tabel 3. Data Testing

No	Kabupaten	Umur 15-19	Umur 20-24	Umur 25-29	Umur 30-34	Umur 35-39	Umur 40-44	Umur 45-49	Kategori
1	Kota Bandar Lampung	9,75	69,55	149,02	121,29	69,8	17,35	1,38	?
2	Kota Metro	10,6	78,7	138,2	109,6	63,6	18,4	1,3	?
3	Lampung	34,5	108,3	135,7	99,3	59,5	15,9	3,3	?

3.2. Perhitungan K-NN

Langkah-langkah proses penghitungan metode K-NN yang menggunakan data latih dan uji dari tabel di atas adalah sebagai berikut.

1. $K=3$ adalah nilai K yang digunakan dalam penelitian ini.
2. Gunakan penghitungan jarak Euclidean untuk menentukan pemisahan antara kumpulan data pelatihan dan pengujian. Perhitungannya adalah sebagai berikut:

$$\sqrt{(49,21 - 9,75)^2 + (134,4 - 69,55)^2 + (138,34 - 149,02)^2 + (102,28 - 121,29)^2 + (58,25 - 69,8)^2 + (15,43 - 17,35)^2 + (3,93 - 1,38)^2} = 79,88521766$$

Tabel 4 merupakan tabel hasil dari perhitungan jarak dengan menggunakan *euclidean distance*.

Tabel 4. Hasil Perhitungan

No	D1	D2	D3
1	79,88521766	68,49222146	30,25928618
2	38,30137204	29,23328411	26,50250743
3	48,77519964	36,7317179	12,2906916
4	67,73790003	53,56755828	23,83500157
5	48,07490406	32,91458643	13,0315003
6	47,70464443	34,98733914	15,13564997
7	62,35794576	48,22456428	13,23905586
8	57,77246836	43,74248164	16,30272063
9	38,99967308	27,23460483	22,74585017
10	24,80339493	18,00805375	30,37301434
11	62,74855457	52,77199541	18,49214698
12	69,33994808	53,55987304	19,2424011
13	58,8063466	52,43739124	27,60959978

3. Setelah menghitung jarak antara data latih (*training*) dan data uji (*testing*), data hasil perhitungan tersebut akan diurutkan untuk menemukan nilai terkecil, seperti yang ditunjukkan pada tabel 5 di bawah ini.

Tabel 5. Urutan Nilai Hasil Perhitungan

No	No D1	D1	No D2	D2	No D3	D3
1	13	79,88521766	13	68,49222146	12	30,25928618
2	2	38,30137204	3	29,23328411	10	26,50250743
3	6	48,77519964	6	36,7317179	1	12,29069160
5	5	48,07490406	4	32,91458643	2	13,03150030
6	4	47,70464443	5	34,98733914	4	15,13564997
7	9	62,35794576	8	48,22456428	3	13,23905586
8	7	57,77246836	7	43,74248164	5	16,30272063
9	3	38,99967308	2	27,23460483	8	22,74585017
10	1	24,80339493	1	18,00805375	13	30,37301434
11	10	62,74855457	10	52,77199541	6	18,49214698

No	No D1	D1	No D2	D2	No D3	D3
12	12	69,33994808	11	53,55987304	7	19,24240110
13	8	58,80634660	9	52,43739124	11	27,60959978

- Langkah berikutnya adalah menentukan nilai $K=3$ dari hasil tabel di atas untuk mengidentifikasi jarak yang paling terdekat sesuai dengan data uji (*testing*) yang ditunjukkan pada tabel 6.

Tabel 6. Urutan terkecil nilai K

No	Data	D1	Kategori	Data	D2	Kategori	Data	D3	Kategori
1	A10	24,803395	Rendah	B10	18,008054	Rendah	C3	12,290692	Rendah
2	A2	38,301372	Sedang	B9	27,234605	Rendah	C5	13,031500	Rendah
3	A9	38,999673	Rendah	B2	29,233284	Sedang	C7	13,239056	Rendah

- Selanjutnya yaitu data akan diurutkan berdasarkan mayoritas yang ada diatas dengan menetapkan kategori berdasarkan hasil prediksi perhitungan secara manual yang telah didapat. Berikut ini adalah tabel 7 yang merupakan target output dari proses yang dilakukan.

Tabel 7. Hasil prediksi data

No	Kabupaten	Umur 15-19	Umur 20-24	Umur 25-29	Umur 30-34	Umur 35-39	Umur 40-44	Umur 45-49	Kategori
1	Kota Bandar Lampung	9,75	69,55	149,02	121,29	69,8	17,35	1,38	Rendah
2	Kota Metro	10,6	78,7	138,2	109,6	63,6	18,4	1,3	Rendah
3	Lampung	34,5	108,3	135,7	99,3	59,5	15,9	3,3	Rendah

3.3. Pengujian

Dalam penelitian ini, digunakan platform pemrosesan data RapidMiner. Berdasarkan penjelasan sebelumnya mengenai langkah-langkah penggunaannya dan output yang dihasilkan, langkah berikutnya adalah menjelaskan validasi data yang dilakukan menggunakan RapidMiner. RapidMiner adalah platform sains data open-source yang digunakan untuk analisis data mining, text mining, dan prediksi [19]. Dengan RapidMiner, pengguna dapat menerapkan teknik deskriptif dan prediktif untuk mendapatkan pemahaman yang lebih mendalam terhadap data yang dimiliki [20]. RapidMiner berfungsi sebagai alat penting dalam menavigasi dan membentuk ulang data, memfasilitasi peneliti dalam pemodelan prediktif, pengelompokan data, dan visualisasi data [21]. Dengan fitur-fiturnya yang ramah pengguna dan mudah beradaptasi, RapidMiner mendukung penelitian ini dalam mengekstraksi wawasan berharga dari kumpulan data yang kompleks, mempercepat proses analisis, dan meningkatkan kualitas pengambilan keputusan melalui basis informasi yang lebih kuat [22]. Tabel 8 berikut ini menunjukkan hasil pengujian menggunakan sistem RapidMiner.

Tabel 8. Matrix evaluasi

Parameter	Nilai
<i>Precision</i>	66.67%
<i>Recall</i>	75.00%
<i>Accuracy</i>	45.00%

Tabel 9. Confusion Matrix

	True Tinggi	True Sedang	True Rendah
Pred. Tinggi	0	1	0
Pred. Sedang	1	0	2
Pred. Rendah	1	2	6

Tabel 9 di menunjukkan hasil pengujian yang diperoleh menggunakan sistem RapidMiner, yang kemudian diinterpretasikan ke dalam kategori berikut:

- Pred.
 - Untuk prediksi kategori tinggi, dari 1 data uji, hasilnya menunjukkan 0 data kategori tinggi, 1 data kategori sedang, 0 data kategori rendah.
 - Untuk prediksi kategori sedang, dari 3 data uji, hasilnya menunjukkan 1 data kategori tinggi, 0 data kategori sedang, 2 data kategori rendah.
 - Untuk kategori rendah, dari 9 data uji, hasilnya menunjukkan 1 data kategori tinggi, 2 data kategori sedang, 6 data kategori rendah.

2. *True*

- a. Untuk prediksi *True* tinggi, dari total 2 data, terdapat 0 data yang tinggi, 1 data yang sedang, 1 data rendah.
- b. Untuk prediksi *True* sedang, dari total 3 data, terdapat 1 data yang tinggi, 0 data yang sedang, 2 data yang rendah.
- c. Untuk prediksi *True* rendah, dari total 8 data, terdapat 0 data yang tinggi, 2 data yang sedang, 6 data yang rendah.

4. **KESIMPULAN**

Dalam penelitian ini, tujuan utamanya adalah untuk meramalkan angka kelahiran berdasarkan kelompok umur ibu menggunakan model K-NN. Namun, hasil analisis menunjukkan bahwa model ini memiliki akurasi yang rendah, yaitu sebesar 45.00%. Hal ini mengindikasikan adanya keterbatasan dalam kemampuan model ini untuk memprediksi angka kelahiran secara tepat. Terdapat beberapa faktor yang dapat mempengaruhi rendahnya akurasi dari model ini. Pertama, variasi karakteristik populasi dapat mempengaruhi hasil prediksi. Setiap kelompok umur ibu mungkin memiliki pola yang berbeda dalam hal angka kelahiran, dan model ini mungkin tidak mampu menangkap perbedaan ini dengan baik. Kedua, distribusi data juga dapat mempengaruhi akurasi model. Jika data tidak terdistribusi dengan baik atau terdapat ketimpangan dalam jumlah sampel untuk setiap kelompok umur ibu, maka model ini mungkin tidak dapat menghasilkan prediksi yang akurat.

Selain itu, jumlah dan jenis fitur yang digunakan dalam model juga dapat berkontribusi terhadap rendahnya akurasi. Jika fitur-fitur yang relevan tidak termasuk dalam model atau jika terlalu sedikit fitur yang digunakan, maka model ini mungkin tidak dapat menangkap kompleksitas yang ada dalam data. Meskipun demikian, mesin pembelajaran K-NN yang diterapkan dalam RapidMiner masih dapat digunakan untuk prediksi angka kelahiran berdasarkan usia ibu. Namun, hasil analisis menunjukkan bahwa masih ada ruang signifikan untuk perbaikan. Penelitian di masa mendatang dapat mempertimbangkan penggunaan metode lain atau meningkatkan kualitas data untuk mencapai akurasi prediksi yang lebih tinggi.

REFERENSI

- [1] D. K. Indahsari and Y. I. Kurniawan, "Aplikasi Prediksi Usia Kelahiran Dengan Metode K-Nearest Neighbor," *J. Kebidanan*, vol. 11, no. 01, p. 1, 2019, doi: 10.35872/jurkeb.v11i01.335.
- [2] Triana, E. Utami, and A. Dwi Hartanto, "INFOKES : Jurnal Ilmiah Rekam Medis dan Informatika Kesehatan IMPLEMENTASI ALGORITMA K-NEAREST NEIGHBOR PADA APLIKASI DETEKSI RESIKO TINGGI PADA KEHAMILAN," vol. 13, no. 2, pp. 64–71, 2023.
- [3] R. Hidayat and T. Astuti, "Diagnosis Preeklamsia pada Ibu Hamil Berdasarkan Algoritme K- Nearest Neighbour," *J. Ilmu-ilmu Inform. dan Manaj. STMIK*, vol. 14, no. 2, pp. 106–116, 2020.
- [4] N. Khoiruzzaman, R. D. Ramadhani, and A. Junaidi, "Hasil Klasifikasi Algoritma Backpropagation dan K-Nearest Neighbor pada Cardiovascular Disease," *J. Dinda Data Sci. Inf. Technol. Data Anal.*, vol. 1, no. 1, pp. 17–27, 2021, doi: 10.20895/dinda.v1i1.141.
- [5] M. Y. Pusadan, A. Ghifari, and Y. Anshori, "Implementasi Data Mining untuk Prediksi Status Proses Persalinan pada Ibu Hamil Menggunakan Algoritma Naive Bayes," *Technomedia J.*, vol. 8, no. 1 Juni, pp. 137–153, 2023, doi: 10.33050/tmj.v8i1.1980.
- [6] H. A. Damar Rani and S. Zuhri, "Sistem Prediksi Kondisi Kelahiran Bayi menggunakan Klasifikasi Naïve Bayes," *Joined J. (Journal Informatics Educ.*, vol. 3, no. 2, p. 48, 2020, doi: 10.31331/joined.v3i2.1432.
- [7] Salsabila, S. Martha, and W. Andani, "Komparasi Algoritma K-Nearest Neighbor Dengan Euclidean Distance Dan Manhattan Distance Untuk Klasifikasi Stunting Balita," *Bul. Ilm. Math. Stat. dan Ter.*, vol. 13, no. 2, pp. 285–292, 2024.
- [8] I. Pembelajaran *et al.*, "JURNAL INFORMATIKA DAN TEKNOLOGI INFORMASI PROGRAM STUDI INFORMATIKA – FAKULTAS TEKNIK - UNIVERSITAS JANABADRA," vol. 8, no. 1, 2023.
- [9] H. I. Purwanto and S. Wibisono, "AHP-CBR Untuk Deteksi Dini Stunting Pada Balita Menggunakan Algoritma Similaritas K-NN," *INFORMATICS Educ. Prof. J. Informatics*, vol. 7, no. 1, pp. 64–73, 2023, [Online]. Available: <https://ejournal-binainsani.ac.id/index.php/ITBI/article/view/2121>
- [10] S. Kasus *et al.*, "Penerapan Data Mining Metode K-Nearest Neighbor Untuk Memprediksi Kelulusan Siswa Sekolah Menengah Pertama," *SMARTICS J.*, vol. 9, no. 1, pp. 14–19, 2023, [Online]. Available: <https://doi.org/10.21067/smartics.v9i1.8088>
- [11] T. Widyanti, S. S. Hilabi, A. Hananto, Tukino, and E. Novalia, "Implementasi K-Means dan K-Nearest Neighbors pada Kategori Siswa Berprestasi," *J. Inf. dan Teknol.*, vol. 5, no. 1, pp. 75–82, 2023, doi: 10.37034/jidt.v5i1.255.
- [12] Juliana Widyastuti Wahyuningsih, "1035325 Studi Deskriptif Sikap dan Pengetahuan Ibu Terhadap Perawatan Tali Pusat Pada Bayi Baru Lahir," *J. Kebidanan J. Med. Sci. Ilmu Kesehat. Akad. Kebidanan*

- Budi Mulia Palembang*, vol. 9, no. 2, pp. 86–89, 2019, doi: 10.35325/kebidanan.v9i2.177.
- [13] M. N. R. Fitriani, B. Priyatna, B. Huda, A. L. Hananto, and T. Tukino, “Implementasi Metode K-Means Untuk Memprediksi Status Kredit Macet,” *J. Sist. Komput. dan Inform.*, vol. 4, no. 3, p. 554, 2023, doi: 10.30865/json.v4i3.5953.
- [14] A. M. M. Anwar, P. Harsani, and A. Maesya, “Penentuan Daerah Prioritas Pelayanan Akta Kelahiran Dengan Metode K-Nn Dan K-Means,” *Komputasi J. Ilm. Ilmu Komput. dan Mat.*, vol. 17, no. 1, pp. 319–328, 2020, doi: 10.33751/komputasi.v17i1.1884.
- [15] U. Nijunniyah and S. S. Hilabi, “Implementation of the K-Nearest Neighbor Algorithm to Predict Sales of Medical Devices in Medical Devices Implementasi Algoritma K-Nearest Neighbor untuk Prediksi Penjualan Alat Kesehatan pada Media Alkes,” vol. 4, no. April, pp. 695–701, 2024.
- [16] D. W. Farkhina Dwi Utari, Amril Mutoi Siregar, “Implementasi Algoritme K-Nearest Neighbor (K-NN) untuk Prediksi Hasil Produksi,” *Sci. Student J. Information, Technol. Sci.*, vol. 1, no. 1, pp. 21–25, 2020.
- [17] R. Rahmadini, Enjel Erika LorencisLubis, Aji Priansyah, Yolanda R.W.N, and Tuti Meutia, “Penerapan Data Mining Untuk Memprediksi Harga Bahan Pangan Di Indonesia Menggunakan Algoritma K-Nearest Neighbor,” *J. Mhs. Akunt. Samudra*, vol. 4, no. 4, pp. 223–235, 2023, doi: 10.33059/jmas.v4i4.7074.
- [18] S. P. Dewi, N. Nurwati, and E. Rahayu, “Penerapan Data Mining Untuk Prediksi Penjualan Produk Terlaris Menggunakan Metode K-Nearest Neighbor,” *Build. Informatics, Technol. Sci.*, vol. 3, no. 4, pp. 639–648, 2022, doi: 10.47065/bits.v3i4.1408.
- [19] D. Damayanti, “Perbandingan Akurasi Software Rapidminer dan Weka Menggunakan Algoritma K-Nearest Neighbor (K-NN),” *J. Syntax Admiration*, vol. 2, no. 6, pp. 994–1006, 2021, doi: 10.46799/jsa.v2i6.247.
- [20] S. Marzukhi, N. Awang, S. N. Alsagoff, and H. Mohamed, “RapidMiner and Machine Learning Techniques for Classifying Aircraft Data,” *J. Phys. Conf. Ser.*, vol. 1997, no. 1, 2021, doi: 10.1088/1742-6596/1997/1/012012.
- [21] R. Puspita and A. Widodo, “Perbandingan Metode K-NN, Decision Tree, dan Naïve Bayes Terhadap Analisis Sentimen Pengguna Layanan BPJS,” *J. Inform. Univ. Pamulang*, vol. 5, no. 4, p. 646, 2021, doi: 10.32493/informatika.v5i4.7622.
- [22] H. Putri, A. I. Purnamasari, A. R. Dikananda, O. Nurdiawan, and S. Anwar, “Penerima Manfaat Bantuan Non Tunai Kartu Keluarga Sejahtera Menggunakan Metode NAÏVE BAYES dan K-NN,” *Build. Informatics, Technol. Sci.*, vol. 3, no. 3, pp. 331–337, 2021, doi: 10.47065/bits.v3i3.1093.