

Klasifikasi Pemilihan

Alat Kontrasepsi pada Program Keluarga Berencana

Menggunakan Algoritma
K-Nearest Neighbor



Klasifikasi Pemilihan

Alat Kontrasepsi pada Program Keluarga Berencana

Menggunakan Algoritma
K-Nearest Neighbor

Euis Sartika | Ibnu Rasyid Munthe
Marnis Nasution | Budiando Bangun

 Penerbit
litrus.

**Klasifikasi Pemilihan Alat Kontrasepsi pada Program Keluarga Berencana
Menggunakan Algoritma K-Nearest Neighbor**

**Penyusun : Euis Sartika
Ibnu Rasyid Munthe
Marnis Nasution
Budianto Bangun**

ISBN : 978-623-127-788-6

Copyright © Januari 2026
Ukuran: 17.6x25 cm; hlm: viii + 62

Hak cipta dilindungi oleh undang-undang. Dilarang mengutip atau memperbanyak baik sebagian ataupun keseluruhan isi buku dengan cara apa pun tanpa izin tertulis dari penerbit.

Desainer sampul : Dicky Gea Nuansa
Penata isi : Bas

Cetakan I: Januari 2026

Diterbitkan, dicetak, dan didistribusikan oleh
CV. Literasi Nusantara Abadi
Perumahan Puncak Joyo Agung Residence Blok B11 Merjosari
Kecamatan Lowokwaru Kota Malang
Telp : +6285887254603, +6285841411519
Email: penerbitlitnus@gmail.com
Web: www.penerbitlitnus.co.id
Anggota IKAPI No. 209/JTI/2018

KATA PENGANTAR

Puji dan syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa atas terselesainya buku monograf berjudul Klasifikasi Pemilihan Alat Kontrasepsi pada Program Keluarga Berencana Menggunakan Algoritma K-Nearest Neighbor (KNN). Buku ini disusun sebagai bentuk pengembangan akademik sekaligus kontribusi ilmiah dari lingkungan Universitas Labuhanbatu, khususnya Fakultas Sains dan Teknologi, Program Studi Sistem Informasi.

Penyusunan buku ini tidak terlepas dari proses pembelajaran, diskusi akademik, serta bimbingan yang konstruktif dari para dosen Universitas Labuhanbatu. Penulis menyampaikan apresiasi dan penghargaan setinggi-tingginya kepada Bapak Ibnu Rasyid Munthe, S.T., M.Kom., Ibu Marnis Nasution, S.Kom., M.Kom., dan Bapak Budianto Bangun, S.Sos., M.Kom., atas arahan, motivasi, serta dukungan intelektual yang diberikan selama proses penelitian dan penulisan buku ini. Kontribusi beliau tidak hanya memperkaya perspektif metodologis, tetapi juga memperkuat landasan konseptual dalam integrasi data mining dengan kajian kependudukan dan kesehatan reproduksi.

Buku ini lahir dari refleksi akademik terhadap pentingnya pemanfaatan data dalam mendukung kebijakan publik, khususnya dalam Program Keluarga Berencana (KB). Dalam konteks transformasi digital, pengolahan data tidak lagi sekadar menjadi aktivitas teknis, melainkan bagian dari strategi peningkatan kualitas layanan publik. Melalui pendekatan machine learning dengan algoritma K-Nearest Neighbor, buku ini berupaya menghadirkan model klasifikasi sebagai alat bantu analitis untuk memahami pola pemilihan alat kontrasepsi pada pasangan usia subur.

Sebagai mahasiswa Universitas Labuhanbatu, penulis menyadari bahwa karya ini merupakan bagian dari proses pembelajaran yang terus berkembang. Oleh karena itu, penulis membuka ruang untuk kritik dan saran yang membangun demi penyempurnaan karya di masa mendatang. Harapannya, buku ini dapat menjadi referensi bagi mahasiswa,

dosen, serta peneliti yang tertarik pada integrasi teknologi informasi, data mining, dan kebijakan kesehatan.

Akhir kata, semoga buku ini memberikan manfaat akademik dan praktis, serta menjadi salah satu kontribusi ilmiah dari Universitas Labuhanbatu dalam pengembangan kajian teknologi dan kependudukan.

Labuhanbatu, 03 Maret 2026

Penulis

DAFTAR ISI

Kata Pengantar.....	iii
Daftar Isi.....	v

BAB 1

Program Keluarga Berencana dalam Dinamika Kependudukan Modern	1
Kependudukan Global dan Pergeseran Paradigma Pasca 2020	1
Program Keluarga Berencana sebagai Strategi Pembangunan	2
Pasangan Usia Subur dan Kompleksitas Pengambilan Keputusan	2
Kesenjangan antara Ketersediaan Data dan Pemanfaatannya.....	2
Urgensi Pendekatan Berbasis Data dalam Program KB.....	3

BAB 2

Kompleksitas Pengambilan Keputusan Kontrasepsi	5
Keputusan Kontrasepsi sebagai Fenomena Multidimensional.....	5
Persepsi Risiko dan Rasionalitas Terbatas.....	6
Pengaruh Faktor Sosial dan Relasional.....	6
Bias Kognitif dalam Keputusan Kesehatan	7
Ketidakpastian Informasi dan Ambiguitas Pilihan.....	7
Implikasi terhadap Program Keluarga Berencana	8

BAB 3

Data sebagai Fondasi Kebijakan Kesehatan Reproduksi	9
Era Kebijakan Berbasis Bukti dan Transformasi Peran Data.....	9
Data Kesehatan Reproduksi sebagai Representasi Realitas Sosial	10

Tantangan Kualitas Data dalam Sistem Pelayanan Publik	10
Dari Data Mentah Menuju Informasi Bermakna.....	11
Data Mining dan Rasionalisasi Kebijakan Program KB.....	11
Risiko Epistemologis dalam Interpretasi Data.....	12

BAB 4

Machine Learning dalam Pelayanan Publik dan Kesehatan.....	13
Transformasi Digital dan Evolusi Pengambilan Keputusan Publik	13
Machine Learning sebagai Paradigma Analitik Modern	14
Relevansi Machine Learning dalam Sistem Kesehatan.....	15
Machine Learning dalam Pelayanan Publik: Potensi dan Paradoks	15
Interpretabilitas dan Rasionalitas Model Prediktif	15
Machine Learning dan Rasionalisasi Keputusan Kontrasepsi.....	16
Batasan Epistemologis Machine Learning dalam Konteks Sosial	16

BAB 5

Algoritma K-Nearest Neighbor dan Logika Kedekatan Data.....	19
Klasifikasi sebagai Pendekatan Epistemologis dalam Analisis Data.....	19
Prinsip Dasar K-Nearest Neighbor: Kedekatan sebagai Rasionalitas.....	19
Ruang Fitur dan Representasi Numerik Realitas Sosial	20
Euclidean Distance dan Makna Matematis Kedekatan	20
Parameter K dan Dilema Stabilitas Model	21
KNN sebagai Algoritma Nonparametrik dan Implikasinya	21
Keterbatasan Konseptual dan Sensitivitas Algoritma.....	22

BAB 6

Desain Penelitian dan Transformasi Data Kontrasepsi.....	23
Desain Penelitian sebagai Kerangka Rasionalitas Analitik	23
Data Akseptor KB sebagai Representasi Fenomena Sosial	25
Seleksi Variabel dan Rasionalitas Representasi	30
Preprocessing Data sebagai Proses Epistemologis	31
Normalisasi dan Stabilitas Struktur Jarak	33
Pembagian Data Latih dan Data Uji sebagai Strategi Generalisasi	34
Risiko Distorsi dan Kesadaran Metodologis.....	38

BAB 7

Implementasi Model KNN Menggunakan Orange Data Mining.....	39
Implementasi Model sebagai Jembatan Konseptual dan Operasional	39
Orange Data Mining dalam Ekosistem Analitik Modern	39
Representasi Data dalam Lingkungan Komputasional	40
Pembentukan Model KNN: Logika Kedekatan dalam Praktik.....	40
Input Data Training.....	41
Pembentukan Model KNN.....	41
Input Data Testing.....	42
Proses Prediksi.....	42
Prediksi dan Inferensi Berbasis Pola Historis.....	43
Hasil Prediksi	43
Evaluasi Model sebagai Proses Validasi Epistemologis.....	44
Sensitivitas Model dan Refleksi Metodologis.....	45

BAB 8

Evaluasi Model dan Interpretasi Kinerja Klasifikasi	47
Evaluasi Model sebagai Proses Epistemologis	47
Keterbatasan Akurasi sebagai Indikator Tunggal.....	48
Precision dan Makna Ketepatan Prediksi.....	48
Recall dan Sensitivitas Model terhadap Variasi Kelas.....	49
F1-Score sebagai Keseimbangan Evaluatif	49
Confusion Matrix dan Struktur Kesalahan Prediksi.....	49
AUC dan Dimensi Diskriminatif Model	51
Matthews Correlation Coefficient dan Stabilitas Prediksi	52
Interpretasi Evaluasi: Antara Statistik dan Realitas Sosial	52

BAB 9

Refleksi, Sintesis, dan Arah Pengembangan Analitik Program Keluarga Berencana	55
Klasifikasi Kontrasepsi sebagai Representasi Realitas Sosial	55
Kontribusi Metodologis Pendekatan Machine Learning.....	56
Batasan Epistemologis dan Risiko Interpretasi.....	56
Implikasi bagi Kebijakan dan Pelayanan Program KB	56
Arah Pengembangan Analitik dan Penelitian Masa Depan.....	57

Penutup Reflektif	57
Daftar Pustaka	59
Profil Penulis.....	61



BAB 1

PROGRAM KELUARGA BERENCANA DALAM DINAMIKA KEPENDUDUKAN MODERN

Kependudukan Global dan Pergeseran Paradigma Pasca 2020

Diskursus kependudukan di Indonesia mengalami transformasi signifikan dalam beberapa tahun terakhir. Jika pada periode sebelumnya perhatian kebijakan lebih berfokus pada pengendalian laju pertumbuhan penduduk, maka pasca 2020 arah pembangunan kependudukan semakin menekankan kualitas penduduk, struktur umur, serta keterkaitannya dengan keberlanjutan pembangunan nasional. Pergeseran ini mencerminkan konsekuensi dari transisi demografi, perubahan sosial, serta tantangan pembangunan yang semakin kompleks [1].

Pertumbuhan penduduk tidak lagi dipandang semata sebagai persoalan kuantitatif, melainkan sebagai fenomena multidimensional yang berkaitan dengan distribusi usia, mobilitas, kesehatan reproduksi, dan kapasitas negara dalam menyediakan layanan dasar. Banyak negara berkembang menghadapi situasi paradoks: di satu sisi menikmati bonus demografi, namun di sisi lain berhadapan dengan tantangan kesiapan institusi sosial dan ekonomi. Dalam konteks ini, kebijakan kependudukan membutuhkan pendekatan yang lebih adaptif, berbasis bukti, dan sensitif terhadap heterogenitas sosial.

Program Keluarga Berencana (KB) menjadi salah satu instrumen kebijakan yang relevan dalam menjawab kompleksitas tersebut. KB tidak lagi dipahami sekadar sebagai strategi pengendalian kelahiran, tetapi sebagai bagian integral dari pembangunan manusia, kesehatan reproduksi, dan stabilitas sosial-ekonomi.

Program Keluarga Berencana sebagai Strategi Pembangunan

Secara konseptual, Program Keluarga Berencana memiliki dimensi strategis yang melampaui fungsi medis atau klinis. KB merupakan intervensi kebijakan publik yang secara langsung berhubungan dengan kualitas hidup keluarga, partisipasi ekonomi perempuan, kesehatan ibu dan anak, serta ketahanan sosial [2]. Oleh karena itu, efektivitas KB tidak dapat diukur hanya melalui indikator penurunan fertilitas, tetapi juga melalui dampaknya terhadap kesejahteraan keluarga dan stabilitas sosial.

Literatur kontemporer menunjukkan bahwa keberhasilan program Keluarga Berencana (KB) dipengaruhi oleh akses informasi, kualitas layanan kesehatan, serta kemampuan sistem dalam memahami karakteristik pengguna layanan. Keputusan penggunaan kontrasepsi merupakan hasil interaksi kompleks antara faktor biologis, sosial, psikologis, ekonomi, dan budaya. Pilihan kontrasepsi tidak selalu ditentukan oleh pertimbangan medis semata, tetapi juga oleh persepsi risiko, norma sosial, dan pengalaman subjektif pengguna [3].

Di negara berkembang, termasuk Indonesia, KB juga berperan dalam menjaga keseimbangan antara pertumbuhan penduduk dan kapasitas pembangunan. Struktur usia yang tidak seimbang dapat memicu tekanan pada sektor pendidikan, kesehatan, dan ketenagakerjaan. Dengan demikian, KB berfungsi sebagai mekanisme stabilisasi demografi sekaligus investasi jangka panjang dalam pembangunan manusia.

Pasangan Usia Subur dan Kompleksitas Pengambilan Keputusan

Penelitian mutakhir menunjukkan bahwa variabel seperti usia, jumlah anak, tingkat pendidikan, dan akses informasi berperan signifikan dalam menentukan kecenderungan penggunaan metode kontrasepsi tertentu. Meski demikian, keberadaan informasi tidak selalu menjamin keputusan yang optimal. Bias kognitif, mispersepsi risiko, serta pengaruh lingkungan sosial sering kali mengarahkan individu pada pilihan yang kurang sesuai secara medis maupun programatik.

Fenomena ini menegaskan bahwa pengambilan keputusan kontrasepsi tidak dapat dipahami hanya melalui pendekatan normatif. Dibutuhkan model analitis yang mampu menangkap pola perilaku berbasis data empiris, sehingga kebijakan dan layanan dapat dirancang lebih responsif terhadap kebutuhan nyata masyarakat.

Kesenjangan antara Ketersediaan Data dan Pemanfaatannya

Ironisnya, di tengah meningkatnya digitalisasi administrasi publik, banyak institusi masih menghadapi kesenjangan antara ketersediaan data dan pemanfaatan analitisnya.

Data akseptor KB yang dikumpulkan secara rutin sering kali berfungsi sebagai arsip administratif, bukan sebagai sumber pengetahuan strategis.

Padahal, data kependudukan dan kesehatan reproduksi menyimpan potensi besar untuk mengungkap pola perilaku, preferensi kontrasepsi, serta faktor-faktor yang memengaruhi keputusan PUS. Tanpa pendekatan analitis yang sistematis, data tersebut tidak memberikan kontribusi optimal terhadap pengambilan keputusan berbasis bukti (evidence-based decision making).

Kemajuan teknologi komputasi dan machine learning membuka peluang baru dalam mengolah data sosial-demografis secara lebih cerdas. Pendekatan ini memungkinkan identifikasi pola yang sulit ditangkap melalui analisis konvensional, sekaligus meningkatkan objektivitas dalam proses rekomendasi layanan.

Urgensi Pendekatan Berbasis Data dalam Program KB

Transformasi digital dalam sektor publik menuntut perubahan paradigma dari kebijakan berbasis asumsi menuju kebijakan berbasis data. Dalam konteks Program Keluarga Berencana, pendekatan berbasis data tidak hanya meningkatkan akurasi analisis, tetapi juga memperkuat legitimasi kebijakan dan kualitas layanan.

Pemanfaatan metode klasifikasi dan machine learning, seperti algoritma K-Nearest Neighbor (KNN), menjadi relevan untuk mengeksplorasi hubungan antara karakteristik PUS dan pilihan kontrasepsi. Model komputasional tidak menggantikan pertimbangan medis, tetapi berfungsi sebagai alat bantu analitis untuk memperkaya proses pengambilan keputusan.

Dengan demikian, integrasi data dan teknologi analitik bukan lagi pilihan, melainkan kebutuhan strategis dalam pengelolaan program kesehatan reproduksi modern. Program Keluarga Berencana dalam dinamika kependudukan modern tidak lagi dapat dipandang sebagai kebijakan pengendalian kelahiran semata. KB merupakan instrumen strategis pembangunan manusia yang berinteraksi dengan struktur demografi, perilaku kesehatan, dan sistem layanan publik. Kompleksitas keputusan kontrasepsi pada tingkat individu menuntut pendekatan yang lebih adaptif dan berbasis data, terutama dalam konteks meningkatnya ketersediaan data administrasi publik.

Pendekatan kependudukan kontemporer menegaskan perlunya transformasi dari analisis deskriptif menuju analisis prediktif. Data akseptor KB memiliki nilai strategis sebagai basis pemodelan perilaku dan pengambilan keputusan berbasis bukti. Integrasi machine learning dalam program KB berpotensi meningkatkan objektivitas, efisiensi,

dan ketepatan rekomendasi layanan tanpa mengabaikan dimensi etika dan konteks sosial.



BAB 2

KOMPLEKSITAS PENGAMBILAN KEPUTUSAN KONTRASEPSI

Keputusan Kontrasepsi sebagai Fenomena Multidimensional

Pemilihan alat kontrasepsi sering kali dipersepsikan sebagai keputusan medis yang rasional dan teknis. Namun dalam realitas sosial, keputusan tersebut merupakan hasil interaksi kompleks antara faktor biologis, psikologis, sosial, ekonomi, dan kultural. Setiap pasangan usia subur (PUS) berada dalam konteks kehidupan yang unik, sehingga pilihan kontrasepsi tidak dapat dipahami semata-mata sebagai respons terhadap rekomendasi klinis.

Jenis alat kontrasepsi yang umum digunakan meliputi kontrasepsi hormonal seperti pil dan suntik, serta kontrasepsi jangka panjang seperti *Intra Uterine Device* (IUD) dan implan. Setiap jenis alat kontrasepsi memiliki karakteristik, kelebihan, dan keterbatasan masing-masing [4].

Pemilihan alat kontrasepsi dipengaruhi oleh berbagai faktor, antara lain usia, jumlah anak, tingkat pendidikan, kondisi kesehatan, serta faktor sosial dan ekonomi. Penelitian terdahulu menunjukkan bahwa perilaku kesehatan dan karakteristik individu memiliki pengaruh signifikan terhadap pilihan metode kontrasepsi yang digunakan [5]. Oleh karena itu, diperlukan pendekatan berbasis data untuk membantu menentukan alat kontrasepsi yang paling sesuai bagi setiap individu.

Kebijakan pemerintah tentang KB saat ini mengarah pada pemakaian kontrasepsi jangka panjang. *Intra Uterine Device* (IUD) adalah salah satu alat kontrasepsi jangka

panjang yang paling efektif dan aman dibandingkan alat kontrasepsi lainnya seperti pil [6].

Persepsi Risiko dan Rasionalitas Terbatas

Salah satu determinan utama dalam pemilihan kontrasepsi adalah persepsi risiko. Risiko dalam konteks ini tidak hanya berkaitan dengan efek samping medis, tetapi juga mencakup kekhawatiran terhadap kenyamanan, stigma sosial, hingga ketakutan terhadap perubahan tubuh. Persepsi risiko bersifat konstruktif, dibentuk oleh informasi yang diterima, pengalaman orang lain, serta narasi yang berkembang di lingkungan sosial.

Konsep bounded rationality menjelaskan bahwa individu mengambil keputusan dengan keterbatasan informasi dan kapasitas kognitif. Dalam situasi ini, pilihan yang diambil bukanlah keputusan optimal secara objektif, melainkan keputusan yang dianggap cukup memuaskan berdasarkan pemahaman yang tersedia. Pada konteks KB, rasionalitas terbatas dapat menyebabkan individu menghindari metode kontrasepsi tertentu bukan karena pertimbangan ilmiah, tetapi karena persepsi yang belum tentu akurat.

Pengaruh Faktor Sosial dan Relasional

Keputusan kontrasepsi pada dasarnya bukan keputusan individual, melainkan keputusan relasional. Hubungan antara suami dan istri, dinamika komunikasi keluarga, serta norma gender berperan besar dalam menentukan pilihan. Pada banyak kasus, keputusan dipengaruhi oleh dominasi salah satu pihak, tekanan keluarga, atau pertimbangan sosial tertentu.



Gambar Diagram Pengaruh Lingkungan Sosial terhadap Preferensi Kontrasepsi

Selain faktor internal keluarga, lingkungan sosial juga memiliki pengaruh signifikan. Norma komunitas, pengalaman kolektif, serta konstruksi budaya mengenai reproduksi

membentuk preferensi kontrasepsi. Fenomena ini menjelaskan mengapa metode tertentu lebih populer di wilayah tertentu meskipun secara medis tersedia berbagai alternatif. Preferensi sosial sering kali membentuk persepsi legitimasi terhadap suatu metode.

Bias Kognitif dalam Keputusan Kesehatan

Psikologi kognitif modern menekankan bahwa manusia rentan terhadap bias dalam pengambilan keputusan. Bias heuristik, seperti availability bias dan confirmation bias, dapat memengaruhi evaluasi individu terhadap metode kontrasepsi. Informasi yang mudah diingat atau pengalaman negatif yang menonjol cenderung memiliki pengaruh lebih besar dibandingkan data statistik yang objektif.

Tabel Contoh Bias Kognitif dalam Keputusan Kesehatan Reproduksi

Jenis Bias	Deskripsi Konseptual	Dampak Potensial dalam Keputusan Kontrasepsi
Availability Bias	Keputusan dipengaruhi oleh informasi yang paling mudah diingat atau paling sering didengar, bukan data objektif.	Memilih metode karena cerita pengalaman orang lain yang menonjol.
Confirmation Bias	Individu cenderung mencari atau mempercayai informasi yang mendukung keyakinan awal.	Mengabaikan informasi medis yang bertentangan dengan persepsi pribadi.
Risk Aversion	Kecenderungan menghindari risiko meskipun alternatif lain lebih rasional secara statistik.	Menghindari metode tertentu karena persepsi risiko efek samping yang diletakkan.

Sebagai contoh, pengalaman efek samping pada satu individu dalam komunitas dapat membentuk persepsi kolektif yang berlebihan terhadap risiko suatu metode. Bias semacam ini memperlihatkan bahwa keputusan kesehatan tidak selalu berbasis bukti ilmiah, melainkan dipengaruhi oleh mekanisme psikologis yang kompleks.

Ketidakpastian Informasi dan Ambiguitas Pilihan

Salah satu tantangan utama dalam keputusan kontrasepsi adalah ketidakpastian informasi. Tidak semua pengguna memiliki akses terhadap konseling yang komprehensif, sementara informasi yang beredar di masyarakat sering kali tidak terstandarisasi. Ambiguitas ini menciptakan kondisi di mana individu harus memilih di antara alternatif yang dipahami secara parsial.

Ketidakpastian informasi berpotensi meningkatkan kecenderungan memilih metode yang paling familiar atau yang direkomendasikan secara informal. Fenomena

ini memperkuat pola keputusan berbasis kebiasaan sosial daripada pertimbangan klinis. Dalam perspektif kebijakan publik, kondisi ini menunjukkan pentingnya sistem pendukung keputusan yang mampu menyediakan informasi objektif berbasis data.

Implikasi terhadap Program Keluarga Berencana

Kompleksitas keputusan kontrasepsi memiliki implikasi langsung terhadap efektivitas program KB. Ketidaksihesuaian antara karakteristik pengguna dan metode yang dipilih dapat memengaruhi keberlanjutan penggunaan, tingkat kepuasan, serta risiko penghentian metode. Dalam skala populasi, akumulasi keputusan individual membentuk pola distribusi penggunaan kontrasepsi yang berdampak pada indikator kependudukan.

Pendekatan berbasis data dan analitik prediktif berpotensi membantu memahami pola keputusan tersebut. Dengan mengidentifikasi hubungan antara variabel sosial-demografis dan pilihan kontrasepsi, model klasifikasi dapat memberikan perspektif tambahan yang lebih sistematis. Hal ini membuka ruang bagi integrasi teknologi dalam mendukung rasionalisasi layanan KB.

Keputusan pemilihan alat kontrasepsi merupakan fenomena multidimensional yang dipengaruhi oleh persepsi risiko, literasi kesehatan, faktor relasional, serta bias kognitif. Rasionalitas individu sering kali dibatasi oleh keterbatasan informasi dan konstruksi sosial yang membentuk preferensi. Kompleksitas ini menjelaskan mengapa keputusan kontrasepsi tidak selalu selaras dengan rekomendasi medis normatif. Dalam konteks program KB, pemahaman terhadap dinamika keputusan menjadi kunci dalam meningkatkan efektivitas layanan.

Faktor yang mempengaruhi peningkatan pengetahuan ibu untuk memilih alat kontrasepsi adalah usia ibu, pendidikan ibu, tenaga kesehatan dan pemberian pendidikan kesehatan [7].



BAB 3

DATA SEBAGAI FONDASI KEBIJAKAN KESEHATAN REPRODUKSI

Era Kebijakan Berbasis Bukti dan Transformasi Peran Data

Tata kelola publik modern mengalami pergeseran dari kebijakan berbasis intuisi menuju kebijakan berbasis bukti (*evidence-based policy*). Perubahan ini dipicu oleh meningkatnya kompleksitas persoalan sosial, kesehatan, dan demografi yang menuntut pendekatan analitik yang lebih sistematis. Dalam konteks ini, data tidak lagi sekadar catatan administratif, melainkan aset strategis yang menentukan kualitas pengambilan keputusan.

Pasca 2020, transformasi digital semakin memperkuat peran data dalam berbagai sektor, termasuk pemerintahan dan layanan kesehatan. Data menjadi instrumen penting untuk memahami pola populasi, memetakan kebutuhan layanan, serta mengevaluasi efektivitas kebijakan. Dalam kesehatan reproduksi, pemanfaatan data menjadi krusial karena kebijakan yang diambil berdampak langsung pada kesejahteraan individu dan keluarga. Tanpa dukungan data yang memadai, kebijakan berisiko tidak sensitif terhadap keragaman karakteristik masyarakat.

Meskipun program KB telah menunjukkan hasil positif dalam menurunkan angka kelahiran, masih terdapat tantangan besar seperti rendahnya personalisasi layanan, keterbatasan tenaga penyuluh, serta ketimpangan akses terhadap informasi dan teknologi, khususnya di wilayah 3T [9].

Data Kesehatan Reproduksi sebagai Representasi Realitas Sosial

Data kesehatan reproduksi memiliki karakteristik unik dibandingkan data administratif lainnya. Variabel yang terkandung di dalamnya tidak hanya mencerminkan kondisi biologis, tetapi juga realitas sosial, budaya, ekonomi, dan perilaku individu. Informasi mengenai usia, jumlah anak, pendidikan, pekerjaan, serta pilihan kontrasepsi merupakan refleksi dari dinamika sosial yang kompleks.

Dalam Program Keluarga Berencana (KB), data akseptor berfungsi sebagai rekaman empiris perilaku reproduksi masyarakat. Setiap entri data bukan sekadar angka, melainkan representasi keputusan, preferensi, dan kondisi sosial yang melatarbelakangi pilihan kontrasepsi. Oleh karena itu, analisis data KB sesungguhnya merupakan proses membaca pola perilaku sosial dalam konteks kesehatan.

Pendekatan ini menegaskan bahwa data kesehatan reproduksi tidak dapat diperlakukan secara mekanistik. Interpretasi data memerlukan pemahaman konseptual mengenai konteks sosial yang membentuk variabel tersebut. Tanpa perspektif ini, analisis berpotensi menghasilkan kesimpulan yang reduksionis.

Tantangan Kualitas Data dalam Sistem Pelayanan Publik

Meskipun data memiliki nilai strategis, kualitas data sering menjadi persoalan mendasar dalam sistem pelayanan publik. Data lapangan rentan terhadap berbagai bentuk ketidaksempurnaan, seperti data hilang (*missing values*), ketidakkonsistenan pencatatan, bias responden, serta noise akibat kesalahan input. Fenomena ini bukan sekadar persoalan teknis, tetapi mencerminkan keterbatasan struktural dalam sistem pengumpulan data.

Dalam konteks KB, variasi kapasitas petugas, perbedaan standar pencatatan, serta kondisi sosial responden memengaruhi reliabilitas data. Sebagai contoh, variabel pekerjaan atau pendidikan dapat mengalami variasi penulisan yang menimbulkan inkonsistensi kategorikal. Ketidaksesuaian ini berdampak langsung pada proses analitik, terutama ketika data digunakan dalam model komputasional.

Tabel Peran Data dalam Siklus Kebijakan Publik Program Keluarga Berencana

Tahapan Siklus	Peran Strategis Data	Makna Analitik	Implikasi Kebijakan
Pengumpulan Data	Menghimpun karakteristik demografis dan sosial pengguna layanan KB.	Membentuk basis representasi empiris populasi.	Menentukan kualitas informasi awal bagi perencanaan program.
Analisis Data	Mengidentifikasi pola, kecenderungan, dan relasi antar variabel.	Menghasilkan wawasan berbasis bukti (<i>evidence-based insight</i>).	Mendukung rasionalitas kebijakan dan alokasi sumber daya.
Pengambilan Keputusan	Menjadi referensi dalam perumusan strategi intervensi dan layanan.	Mengurangi subjektivitas interpretasi kebijakan.	Memperkuat konsistensi kebijakan berbasis data.
Evaluasi Kebijakan	Menilai efektivitas intervensi dan perubahan pola populasi.	Memungkinkan refleksi dan penyesuaian kebijakan adaptif.	Mendukung keberlanjutan dan akuntabilitas program KB.

Literatur kontemporer menegaskan bahwa kualitas data merupakan determinan utama performa model machine learning. Algoritma yang canggih sekalipun tidak mampu menghasilkan prediksi yang valid jika data input mengandung distorsi sistematis. Dengan demikian, preprocessing data bukan sekadar prosedur teknis, tetapi proses epistemologis dalam memastikan validitas representasi realitas.

Dari Data Mentah Menuju Informasi Bermakna

Transformasi data menjadi informasi merupakan proses konseptual yang melibatkan seleksi, pembersihan, dan interpretasi. Data mentah pada dasarnya bersifat pasif; makna hanya muncul melalui proses analitik yang sistematis. Dalam kebijakan kesehatan reproduksi, kemampuan mengubah data menjadi wawasan strategis menentukan kualitas intervensi yang dirumuskan.

Analisis data memungkinkan identifikasi pola yang tidak mudah diamati secara intuitif. Hubungan antara variabel sosial-demografis dan pilihan kontrasepsi, misalnya, dapat mengungkap kecenderungan perilaku yang relevan bagi perencanaan layanan. Pendekatan ini memperlihatkan bahwa data berfungsi sebagai instrumen reduksi ketidakpastian dalam pengambilan keputusan.

Namun demikian, proses transformasi data harus disertai kesadaran metodologis. Tidak semua pola statistik memiliki makna kausal atau normatif. Distingui antara korelasi dan interpretasi kebijakan menjadi aspek penting dalam menjaga integritas analisis.

Data Mining dan Rasionalisasi Kebijakan Program KB

Pemanfaatan data mining dalam kebijakan publik semakin relevan seiring meningkatnya volume dan kompleksitas data sosial. Teknik klasifikasi, klusterisasi, dan prediksi memungkinkan ekstraksi pengetahuan tersembunyi (hidden patterns) yang berpotensi mendukung rasionalisasi kebijakan. Dalam Program KB, data mining membuka ruang untuk memahami struktur preferensi kontrasepsi secara lebih sistematis.

Pendekatan komputasional tidak dimaksudkan untuk menggantikan kebijakan normatif, tetapi memperkaya basis pertimbangan keputusan. Model klasifikasi, seperti K-Nearest Neighbor (KNN), dapat membantu mengidentifikasi kemiripan karakteristik pengguna dan kecenderungan pilihan kontrasepsi. Informasi ini berpotensi menjadi alat bantu dalam proses konseling maupun perencanaan distribusi layanan.

Diskursus modern mengenai data-driven policy menegaskan bahwa integrasi analisis data dan kebijakan publik bukan lagi pilihan, melainkan kebutuhan struktural.

Dalam konteks KB, penggunaan analitik prediktif berpotensi meningkatkan objektivitas, konsistensi, dan efisiensi layanan. Dengan ini *Knowledge* yang diperoleh dari hasil mining terhadap sekumpulan data tersebut dapat membantu dalam pengambilan sebuah keputusan [8].

Risiko Epistemologis dalam Interpretasi Data

Meskipun data menawarkan peluang rasionalisasi, interpretasi data tidak bebas dari risiko epistemologis. Bias pengukuran, ketidakseimbangan data, serta asumsi metodologis dapat memengaruhi kesimpulan analisis. Dalam model klasifikasi, misalnya, distribusi kelas yang tidak seimbang dapat menciptakan ilusi performa tinggi pada kelas dominan.

Selain itu, data sosial-demografis tidak selalu merepresentasikan determinan keputusan secara lengkap. Faktor psikologis, nilai budaya, dan dinamika relasional sering kali tidak terekam dalam dataset kuantitatif. Kesadaran terhadap keterbatasan ini penting untuk menghindari overgeneralisasi hasil analisis.

Pendekatan reflektif terhadap data menegaskan bahwa analisis kuantitatif harus dipahami sebagai representasi parsial realitas. Integrasi perspektif teoritis dan kontekstual menjadi prasyarat dalam menjaga validitas konseptual.

Data dalam kebijakan kesehatan reproduksi berfungsi sebagai fondasi rasionalitas keputusan publik modern. Dalam Program Keluarga Berencana, data akseptor merepresentasikan dinamika sosial-demografis yang kompleks dan berpotensi menghasilkan wawasan strategis melalui analisis sistematis. Namun, kualitas data dan kesadaran metodologis menjadi faktor krusial dalam memastikan validitas interpretasi. Pemanfaatan data mining membuka peluang rasionalisasi layanan KB, sekaligus menuntut refleksi kritis terhadap keterbatasan representasi data.

Data kesehatan reproduksi harus dipahami sebagai konstruksi sosial yang memerlukan interpretasi kontekstual, bukan sekadar objek numerik. Transformasi data menjadi informasi bermakna merupakan proses epistemologis yang menentukan kualitas kebijakan. Integrasi analitik komputasional dalam Program KB mencerminkan pergeseran paradigma menuju kebijakan berbasis bukti. Kesadaran terhadap bias dan keterbatasan data menjadi landasan penting dalam menjaga integritas analisis.



BAB 4

MACHINE LEARNING DALAM PELAYANAN PUBLIK DAN KESEHATAN

Transformasi Digital dan Evolusi Pengambilan Keputusan Publik

Transformasi digital pasca 2020 telah mengubah cara institusi publik memproduksi, mengelola, dan memanfaatkan informasi. Layanan publik tidak lagi hanya berorientasi pada prosedur administratif, tetapi semakin menekankan efektivitas, presisi, dan responsivitas berbasis data. Dalam konteks ini, teknologi komputasional berkembang dari sekadar alat otomasi menjadi instrumen strategis dalam mendukung pengambilan keputusan.

Kompleksitas persoalan sosial dan kesehatan modern menuntut pendekatan analitik yang mampu mengolah data dalam skala besar dan variabilitas tinggi. Model-model tradisional berbasis statistik deskriptif sering kali tidak memadai untuk menjelaskan pola dinamis yang terbentuk dari interaksi variabel sosial. Machine learning muncul sebagai pendekatan yang menawarkan kemampuan adaptif dalam mengenali struktur data dan menghasilkan prediksi berbasis pola empiris.

Perubahan ini menandai pergeseran epistemologis dalam tata kelola publik: keputusan tidak lagi semata-mata berbasis pengalaman atau intuisi, tetapi increasingly berbasis model analitik.

Machine Learning sebagai Paradigma Analitik Modern

Machine learning (ML) merupakan cabang kecerdasan buatan yang berfokus pada pengembangan sistem yang mampu belajar dari data tanpa pemrograman eksplisit berbasis aturan tetap. Paradigma ini memungkinkan algoritma mengenali pola, hubungan, dan struktur dalam dataset melalui proses pelatihan (training). Berbeda dari pendekatan deterministik klasik, ML bekerja dengan prinsip generalisasi berbasis pengalaman data.

Tabel Perbandingan Pendekatan Statistik Tradisional vs Machine Learning

Aspek Perbandingan	Statistik Tradisional	Machine Learning
Tujuan Analisis	Menguji hipotesis, menjelaskan hubungan antar variabel	Prediksi dan klasifikasi berbasis pola data
Pendekatan Model	Berbasis asumsi dan model matematis eksplisit	Berbasis data, sering tanpa asumsi distribusi ketat
Ketergantungan Asumsi	Tinggi (normalitas, linearitas, dll.)	Relatif rendah, lebih fleksibel
Interpretabilitas	Umumnya mudah dijelaskan	Bisa kompleks tergantung algoritma
Kebutuhan Data	Dapat bekerja dengan data kecil	Biasanya membutuhkan data lebih besar
Penanganan Nonlinearitas	Terbatas	Sangat baik
Fokus Analisis	Penjelasan & inferensi	Prediksi & performa model
Contoh Metode	Regresi linear, uji-t, ANOVA	KNN, Decision Tree, Neural Network

Dalam perspektif konseptual, ML merepresentasikan pendekatan probabilistik terhadap realitas. Model tidak menyatakan kepastian mutlak, tetapi estimasi berbasis kecenderungan historis. Hal ini menjadikan machine learning sangat relevan dalam domain sosial dan kesehatan, di mana variabilitas perilaku manusia sulit direduksi ke aturan tunggal.

Supervised learning, sebagai salah satu pendekatan utama ML, menggunakan data berlabel untuk membangun model klasifikasi atau prediksi. Dalam konteks Program Keluarga Berencana (KB), pendekatan ini memungkinkan sistem mempelajari hubungan antara karakteristik pasangan usia subur (PUS) dan pilihan kontrasepsi yang diambil.

Relevansi Machine Learning dalam Sistem Kesehatan

Sektor kesehatan merupakan salah satu domain yang paling diuntungkan oleh perkembangan machine learning. Variasi data klinis, demografis, dan perilaku pasien menciptakan kebutuhan akan sistem analitik yang mampu mengidentifikasi pola kompleks secara efisien. Machine learning telah digunakan dalam berbagai aplikasi, mulai dari prediksi penyakit, analisis citra medis, hingga rekomendasi intervensi kesehatan.

Keunggulan ML dalam kesehatan terletak pada kemampuannya menangani multidimensionalitas data. Faktor biologis, sosial, dan lingkungan dapat dianalisis secara simultan tanpa asumsi linearitas yang ketat. Pendekatan ini memperluas kapasitas sistem kesehatan dalam memahami determinan perilaku dan risiko populasi.

Dalam konteks kesehatan reproduksi, ML menawarkan peluang untuk memahami pola keputusan kontrasepsi yang dipengaruhi oleh interaksi variabel sosial-demografis. Pendekatan ini sejalan dengan kebutuhan kebijakan modern yang menuntut presisi dan sensitivitas terhadap heterogenitas populasi.

Machine Learning dalam Pelayanan Publik: Potensi dan Paradoks

Adopsi machine learning dalam pelayanan publik menghadirkan peluang sekaligus paradoks. Di satu sisi, ML memungkinkan peningkatan efisiensi, konsistensi, dan objektivitas keputusan. Di sisi lain, penggunaan algoritma dalam kebijakan publik memunculkan pertanyaan etis dan epistemologis mengenai transparansi, akuntabilitas, dan bias sistemik.

Ada tiga metode utama dalam ML, yaitu supervised learning, yang memanfaatkan data dengan label untuk membuat model prediktif, unsupervised learning, yang digunakan untuk menemukan pola atau struktur dari data yang belum diberi label, serta reinforcement learning, yang memungkinkan sistem belajar dari pengalaman dan mendapatkan umpan balik dari lingkungannya[10].

Dalam Program KB, risiko bias dapat muncul jika dataset tidak merepresentasikan keragaman karakteristik PUS secara proporsional. Oleh karena itu, integrasi ML dalam pelayanan publik menuntut kesadaran metodologis dan etis yang kuat.

Interpretabilitas dan Rasionalitas Model Prediktif

Salah satu tantangan utama dalam machine learning adalah interpretabilitas model. Tidak semua algoritma menghasilkan struktur keputusan yang mudah dipahami

manusia. Dalam domain kesehatan dan kebijakan publik, transparansi menjadi faktor krusial karena keputusan berdampak langsung pada individu.

Model yang interpretabel memungkinkan pengguna memahami dasar prediksi yang dihasilkan sistem. Hal ini penting untuk menjaga kepercayaan institusional dan legitimasi penggunaan teknologi. Algoritma sederhana seperti K-Nearest Neighbor (KNN) sering dipandang lebih transparan dibanding model kompleks berbasis jaringan saraf dalam (deep learning), karena logika klasifikasinya berbasis kedekatan data. Interpretabilitas bukan sekadar aspek teknis, tetapi elemen normatif dalam penerapan machine learning yang bertanggung jawab.

Machine Learning dan Rasionalisasi Keputusan Kontrasepsi

Keputusan kontrasepsi dalam Program KB mencerminkan dinamika sosial-demografis yang kompleks, dipengaruhi oleh usia, pendidikan, jumlah anak, dan pekerjaan. Pola tersebut tidak selalu mudah dikenali secara manual, sehingga machine learning menawarkan pendekatan sistematis melalui model klasifikasi berbasis data historis.

Klasifikasi pada data yang berjumlah besar dan kompleks jika dilakukan secara manual akan sulit dan membutuhkan waktu yang lama. Hal ini dapat diatasi dengan pembangunan suatu tools klasifikasi karena adanya tools klasifikasi data yang berjumlah besar dan kompleks tersebut akan lebih mudah dan cepat dikelompokkan sesuai dengan kelasnya masing-masing[11].

Model ini tidak menentukan pilihan individu, tetapi menyediakan estimasi pola yang dapat menjadi referensi tambahan dalam layanan KB. Pendekatan berbasis ML membantu meningkatkan konsistensi rekomendasi, mengurangi subjektivitas, dan mengoptimalkan pemanfaatan data akseptor.

Batasan Epistemologis Machine Learning dalam Konteks Sosial

Meskipun menjanjikan, machine learning memiliki batasan epistemologis yang penting untuk disadari. Model ML bekerja berdasarkan pola statistik, bukan pemahaman kausalitas sosial secara menyeluruh. Data sosial tidak selalu menangkap dimensi psikologis, budaya, dan relasional yang memengaruhi keputusan individu.

Selain itu, prediksi berbasis data historis tidak selalu relevan terhadap perubahan konteks sosial. Preferensi kontrasepsi dapat berubah akibat faktor kebijakan, edukasi, atau dinamika budaya. Kesadaran terhadap keterbatasan ini penting untuk menghindari determinisme algoritmik. Machine learning harus diposisikan sebagai alat bantu analitik, bukan otoritas keputusan absolut.

Transformasi digital memperluas peran machine learning sebagai paradigma analitik modern dalam pelayanan publik dan kesehatan. Machine learning menawarkan kemampuan adaptif dalam mengenali pola data kompleks, termasuk dalam konteks kesehatan reproduksi dan Program Keluarga Berencana. Meskipun demikian, penerapannya menghadirkan tantangan terkait kualitas data, bias algoritmik, dan interpretabilitas model. ML berfungsi sebagai instrumen rasionalisasi keputusan berbasis pola empiris, bukan pengganti pertimbangan profesional.

Integrasi machine learning dalam Program KB mencerminkan pergeseran paradigma menuju kebijakan dan layanan berbasis data. Interpretabilitas model menjadi elemen normatif penting dalam konteks kesehatan. Kesadaran terhadap bias dan keterbatasan data sosial menjadi prasyarat dalam penggunaan ML yang bertanggung jawab. Machine learning memperluas kapasitas analitik institusi tanpa menghilangkan peran manusia dalam pengambilan keputusan.



BAB 5

ALGORITMA K-NEAREST NEIGHBOR DAN LOGIKA KEDEKATAN DATA

Klasifikasi sebagai Pendekatan Epistemologis dalam Analisis Data

Klasifikasi dalam machine learning bukan sekadar prosedur teknis, melainkan pendekatan epistemologis dalam memahami struktur realitas melalui data. Ketika suatu fenomena sosial atau kesehatan direpresentasikan dalam bentuk variabel, proses klasifikasi berfungsi untuk mengidentifikasi pola keterkaitan antaratribut dan kategori hasil. Dengan demikian, algoritma klasifikasi bekerja sebagai instrumen inferensi berbasis kemiripan karakteristik.

Dalam konteks pemilihan alat kontrasepsi pada Program Keluarga Berencana (KB), klasifikasi memungkinkan peneliti membaca kecenderungan keputusan berdasarkan karakteristik pasangan usia subur (PUS). Variabel sosial-demografis yang tampak sederhana sesungguhnya menyimpan relasi kompleks yang sulit diamati melalui pendekatan manual. Algoritma klasifikasi berperan sebagai mekanisme sistematis dalam mengekstraksi struktur tersebut.

Di antara berbagai metode klasifikasi, K-Nearest Neighbor (KNN) menempati posisi unik karena kesederhanaan logika dan transparansi konseptualnya.

Prinsip Dasar K-Nearest Neighbor: Kedekatan sebagai Rasionalitas

K-Nearest Neighbor beroperasi berdasarkan prinsip fundamental bahwa objek dengan karakteristik serupa cenderung berada dalam kategori yang sama. Logika ini berakar

pada konsep similarity, yaitu gagasan bahwa kemiripan atribut dapat digunakan sebagai dasar penentuan kelas. Berbeda dari algoritma berbasis model parametrik, KNN tidak membangun fungsi prediktif eksplisit, melainkan mengandalkan struktur kedekatan dalam ruang fitur (feature space).

Dalam kerangka ini, setiap data direpresentasikan sebagai titik dalam ruang multidimensi, di mana setiap dimensi merefleksikan variabel tertentu. Proses klasifikasi dilakukan dengan menghitung jarak antara data baru dan data historis, kemudian menentukan kelas berdasarkan mayoritas tetangga terdekat. Pendekatan ini menjadikan KNN sebagai algoritma berbasis instance (instance-based learning), karena keputusan ditentukan langsung oleh data latih.

Konsep kedekatan dalam KNN tidak sekadar matematis, tetapi mencerminkan asumsi konseptual bahwa kemiripan kuantitatif memiliki relevansi terhadap kategori fenomena.

Ruang Fitur dan Representasi Numerik Realitas Sosial

Agar logika kedekatan dapat dioperasionalkan, seluruh atribut dalam dataset harus direpresentasikan dalam bentuk numerik. Transformasi ini merupakan langkah konseptual penting, terutama ketika variabel yang dianalisis berasal dari domain sosial. Variabel seperti pendidikan dan pekerjaan pada dasarnya bersifat kategorikal, namun dalam KNN perlu diterjemahkan ke bentuk angka agar dapat dihitung jaraknya.

Proses transformasi tersebut membawa implikasi metodologis. Representasi numerik tidak selalu mencerminkan jarak konseptual yang sebenarnya. Misalnya, pengkodean kategori pekerjaan ke dalam skala ordinal dapat menciptakan persepsi jarak matematis yang tidak identik dengan jarak sosial. Oleh karena itu, interpretasi hasil KNN harus mempertimbangkan bahwa jarak komputasional merupakan konstruksi analitik, bukan representasi absolut realitas.

Dalam analisis pemilihan kontrasepsi, ruang fitur menjadi arena di mana variabel sosial-demografis dipetakan sebagai koordinat numerik yang menentukan struktur kedekatan antarindividu.

Euclidean Distance dan Makna Matematis Kedekatan

Euclidean Distance merupakan metrik jarak paling umum dalam KNN. Secara geometris, jarak Euclidean mengukur kedekatan dua titik berdasarkan perbedaan kuadrat pada setiap dimensi. Dalam interpretasi konseptual, jarak ini merepresentasikan tingkat kemiripan keseluruhan atribut.

Penggunaan Euclidean Distance mengandung asumsi bahwa seluruh variabel memiliki kontribusi proporsional terhadap struktur jarak. Namun, asumsi ini hanya valid jika skala variabel seragam. Variabel dengan rentang nilai besar berpotensi mendominasi perhitungan jarak, sehingga normalisasi menjadi prosedur krusial dalam menjaga keseimbangan kontribusi fitur.

Dengan demikian, jarak dalam KNN tidak sekadar hasil perhitungan matematis, tetapi refleksi dari struktur representasi data yang telah melalui transformasi metodologis tertentu.

Parameter K dan Dilema Stabilitas Model

Nilai K dalam KNN menentukan jumlah tetangga yang digunakan dalam proses klasifikasi. Parameter ini memiliki implikasi langsung terhadap sensitivitas dan stabilitas model. Nilai K kecil cenderung menghasilkan model yang responsif terhadap variasi lokal, tetapi rentan terhadap noise. Sebaliknya, nilai K besar menghasilkan generalisasi lebih stabil, namun berisiko mengaburkan batas antar kelas.

Dilema ini mencerminkan trade-off klasik antara bias dan variance dalam machine learning. Tidak terdapat nilai K universal yang optimal; pemilihan parameter harus mempertimbangkan karakteristik dataset dan tujuan analisis. Dalam konteks KB, pemilihan nilai K berkaitan dengan keseimbangan antara sensitivitas terhadap variasi individu dan konsistensi prediksi populasi.

Keputusan parameter K pada akhirnya merupakan keputusan epistemologis mengenai seberapa luas konteks kedekatan dianggap relevan dalam menentukan kategori.

KNN sebagai Algoritma Nonparametrik dan Implikasinya

KNN dikategorikan sebagai algoritma nonparametrik karena tidak mengasumsikan bentuk distribusi data tertentu. Fleksibilitas ini menjadikannya adaptif terhadap berbagai struktur dataset, termasuk data sosial-demografis yang heterogen. Namun, fleksibilitas tersebut juga membawa konsekuensi komputasional, karena proses klasifikasi memerlukan perhitungan jarak terhadap seluruh data latih.

Dalam dataset berukuran kecil hingga menengah, pendekatan ini relatif efisien. Tetapi pada skala besar, kompleksitas komputasi meningkat signifikan. Oleh karena itu, penggunaan KNN harus mempertimbangkan keseimbangan antara kesederhanaan konsep dan efisiensi operasional.

Dalam penelitian pemilihan kontrasepsi, KNN menawarkan keuntungan interpretabilitas yang tinggi karena logika klasifikasinya mudah dijelaskan.

Keterbatasan Konseptual dan Sensitivitas Algoritma

Meskipun sederhana, KNN memiliki sensitivitas metodologis yang signifikan. Algoritma ini sangat dipengaruhi oleh skala data, distribusi kelas, serta representasi numerik variabel kategorikal. Ketidakseimbangan kelas (*class imbalance*) dapat menyebabkan bias prediksi terhadap kelas dominan. Selain itu, korelasi antarvariabel dapat memengaruhi struktur jarak secara tidak proporsional.

Kesadaran terhadap keterbatasan ini penting untuk menjaga validitas interpretasi. KNN tidak menghasilkan model kausal, melainkan model berbasis kemiripan statistik. Oleh karena itu, hasil klasifikasi harus dipahami sebagai estimasi kecenderungan, bukan determinasi absolut.

Algoritma K-Nearest Neighbor berlandaskan prinsip kedekatan dan kemiripan data sebagai dasar klasifikasi. Representasi numerik variabel sosial-demografis memungkinkan operasionalisasi konsep *similarity* dalam ruang fitur multidimensi. Euclidean Distance berfungsi sebagai metrik kedekatan matematis yang sensitif terhadap skala data. Parameter K mencerminkan dilema antara sensitivitas lokal dan stabilitas global. KNN menawarkan kesederhanaan konseptual dan interpretabilitas tinggi, namun rentan terhadap distorsi representasi dan distribusi data.

KNN merepresentasikan paradigma klasifikasi berbasis kedekatan yang relevan untuk analisis sosial-demografis dalam Program KB. Transformasi data kategorikal ke numerik bukan sekadar prosedur teknis, tetapi konstruksi epistemologis yang memengaruhi makna jarak. Pemilihan parameter dan metrik jarak menentukan struktur inferensi model. Kesadaran terhadap keterbatasan algoritma menjadi prasyarat dalam menjaga integritas analisis prediktif.



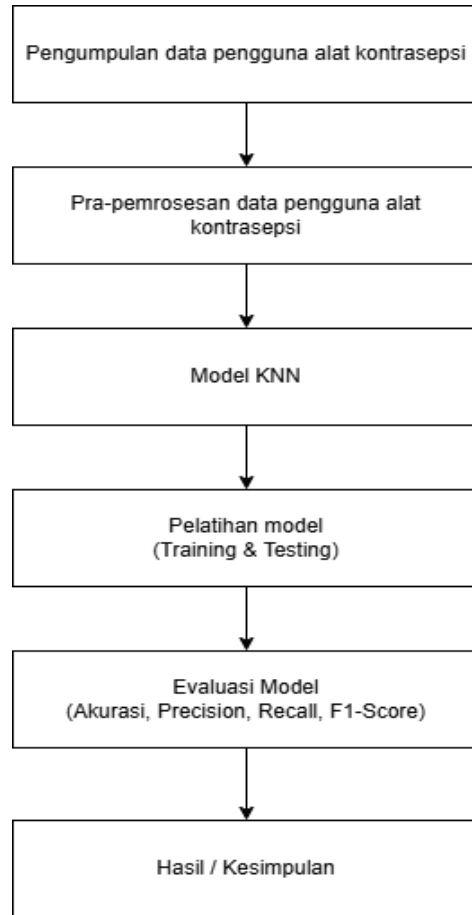
BAB 6

DESAIN PENELITIAN DAN TRANSFORMASI DATA KONTRASEPSI

Desain Penelitian sebagai Kerangka Rasionalitas Analitik

Desain penelitian dalam studi berbasis machine learning tidak hanya berfungsi sebagai prosedur metodologis, tetapi sebagai kerangka rasionalitas yang menentukan validitas inferensi analitik. Dalam penelitian klasifikasi pemilihan alat kontrasepsi, desain penelitian berperan penting dalam memastikan bahwa data, variabel, dan proses transformasi selaras dengan tujuan analisis.

Berbeda dari pendekatan statistik konvensional yang menitikberatkan pada pengujian hipotesis, pendekatan machine learning berorientasi pada pembentukan model prediktif berbasis pola. Konsekuensinya, kualitas desain penelitian sangat ditentukan oleh representasi variabel, struktur dataset, serta keandalan proses preprocessing. Kesalahan pada tahap awal dapat menghasilkan bias sistematis yang sulit diperbaiki pada tahap pemodelan.



Gambar 1 Kerangka Penelitian

Penjelasan tahapan:

1. Pengumpulan data pengguna alat kontrasepsi
Mengumpulkan data pasangan/pengguna alat kontrasepsi dari rekam medis, survei, atau laporan program KB. Variabelnya bisa berupa usia, jumlah anak, pendidikan, pekerjaan, riwayat kesehatan, jenis alat kontrasepsi yang digunakan, dan faktor lain yang relevan.
2. Pra-pemrosesan data pengguna alat kontrasepsi
Data mentah dibersihkan dan disiapkan. Kejadiannya: menghapus data ganda, menangani data hilang (missing value), mengubah kategori ke bentuk numerik, dan menormalisasi data agar perhitungan jarak di KNN tidak berat sebelah. Hasil tahap ini adalah dataset rapi yang siap dianalisis.

3. Model KNN

Setelah karakter data jelas, kemudian menetapkan algoritma utama yang dipakai, yaitu K-Nearest Neighbor (KNN). Di tahap ini juga dijelaskan alasan pemilihan KNN (sederhana, cocok untuk klasifikasi, mudah diimplementasikan di Orange) dan gambaran awal parameter yang akan dicoba, seperti nilai k dan jenis jarak (distance metric).

4. Pelatihan model (Training & Testing)

Dataset dibagi menjadi data latih dan data uji. Model KNN dilatih menggunakan data latih untuk mengenali pola hubungan antara fitur (profil pengguna) dan kelas (jenis alat kontrasepsi). Lalu model diuji dengan data uji untuk melihat kemampuan prediksi pada data yang belum pernah “dilihat” model sebelumnya.

5. Evaluasi model (Akurasi, Precision, Recall, F1-Score)

Kinerja KNN diukur dengan metrik evaluasi:

Akurasi: seberapa besar prediksi yang tepat.

Precision: ketepatan model ketika menyatakan satu jenis kontrasepsi.

Recall: kemampuan model menangkap semua kasus pada satu kelas.

F1-Score: keseimbangan antara precision dan recall.

6. Hasil / Kesimpulan

Bagian akhir berisi kesimpulan penelitian dan rekomendasi. Merangkum apakah model KNN layak dijadikan alat bantu keputusan bagi tenaga kesehatan, apa keterbatasan penelitian, dan apa saran untuk pengembangan model atau penelitian lanjutan. Tahap ini menutup alur kerja yang digambarkan pada flowchart.

Dalam konteks Program Keluarga Berencana (KB), desain penelitian perlu mempertimbangkan karakteristik data sosial-demografis yang dinamis, heterogen, dan kontekstual.

Data Akseptor KB sebagai Representasi Fenomena Sosial

Data yang digunakan dalam penelitian ini merupakan data pasangan usia subur (PUS) peserta Program Keluarga Berencana (KB) di Dinas Pengendalian Penduduk dan Keluarga Berencana (DPPKB) Kabupaten Labuhanbatu. Data tersebut memuat informasi karakteristik pasangan yang berkaitan dengan pemilihan alat kontrasepsi. Langkah awal dalam proses Knowledge Discovery in Database (KDD) adalah seleksi data. Pada tahap ini, data dipilih dan disusun menggunakan aplikasi Microsoft Excel untuk memastikan data yang digunakan sesuai dengan kebutuhan penelitian.

Penting untuk disadari bahwa data lapangan pada domain sosial tidak pernah sepenuhnya netral. Data merupakan hasil konstruksi sistem pencatatan, interaksi petugas-responden, serta interpretasi kategori administratif. Oleh karena itu, analisis data KB menuntut kehati-hatian epistemologis dalam memahami bahwa dataset adalah representasi parsial realitas, bukan refleksi absolut. Pendekatan ini menempatkan desain penelitian sebagai jembatan antara fenomena sosial dan model komputasional.

Tabel 1 Dataset Penelitian

ID	Usia Suami	Usia Istri	Pekerjaan Suami	Pekerjaan Istri	Pendidikan Suami	Pendidikan Istri	Jumlah Anak	Usia Anak Terakhir (Bulan)	Pilihan KB
1	38	32	Petani	Swasta	SMP	SMP	2	14	Suntik
2	39	32	PNS	IRT	S1	D3	2	20	Suntik
3	42	38	Dagang	PNS	SMP	SMP	4	35	Implant
4	43	38	Buruh	IRT	D3	SMA	4	24	Implant
5	26	20	Petani	Dagang	S1	SD	1	10	Pil
6	45	39	Buruh	Dagang	SMP	SMP	3	35	Implant
7	43	39	Wiraswasta	Swasta	SD	SD	3	27	Implant
8	36	33	Swasta	IRT	SMP	SMA	2	20	Suntik
9	42	36	Petani	IRT	SMP	D3	4	33	Implant
10	48	44	Dagang	PNS	SD	SMP	6	60	IUD
11	52	44	Petani	PNS	SD	SD	4	40	IUD
12	31	25	Petani	Swasta	S1	SMA	1	7	Pil
13	28	23	Petani	IRT	S1	SMA	0	0	Kondom
14	26	21	Petani	PNS	D3	SMA	1	9	Pil
15	38	31	Buruh	Swasta	SD	SMA	3	20	Suntik
16	44	40	PNS	Dagang	SD	SMA	3	35	Implant
17	33	27	Dagang	PNS	SD	D3	2	12	Pil
18	37	29	Dagang	PNS	S1	SMP	2	14	Pil
19	41	33	Wiraswasta	Swasta	SMA	D3	3	16	Suntik
20	31	26	Buruh	Dagang	SMA	SMP	1	12	Pil
21	51	43	Petani	PNS	SMA	D3	4	36	IUD
22	29	26	Wiraswasta	Dagang	SMP	S1	2	11	Pil
23	49	43	Swasta	Dagang	SMP	SD	4	60	IUD
24	41	38	Buruh	PNS	SMA	SMA	4	27	Implant
25	51	45	Swasta	Swasta	SMA	D3	5	40	IUD
26	42	34	PNS	PNS	D3	S1	2	18	Suntik
27	34	28	Wiraswasta	PNS	SMP	D3	2	11	Pil
28	41	37	Swasta	IRT	SD	SMA	4	24	Implant
29	46	43	Wiraswasta	Dagang	SD	S1	5	48	IUD
30	31	24	Buruh	PNS	SD	SD	1	10	Pil
31	49	44	Petani	PNS	SD	D3	4	60	IUD
32	39	35	Dagang	Dagang	SMA	SD	2	22	Suntik

ID	Usia Suami	Usia Istri	Pekerjaan Suami	Pekerjaan Istri	Pendidikan Suami	Pendidikan Istri	Jumlah Anak	Usia Anak Terakhir (Bulan)	Pilihan KB
33	39	31	Wiraswasta	IRT	SD	SMP	2	16	Suntik
34	28	21	PNS	PNS	SMP	SMA	0	0	Kondom
35	51	44	Petani	Dagang	SD	SMA	6	48	IUD
36	33	25	Buruh	Swasta	D3	S1	0	0	Kondom
37	44	41	Petani	PNS	S1	SD	4	60	IUD
38	39	34	Dagang	Dagang	S1	SD	3	14	Suntik
39	29	25	Buruh	Dagang	S1	SMA	1	6	Pil
40	48	45	Buruh	Dagang	SMP	S1	6	44	IUD
41	46	41	Buruh	PNS	SMA	S1	4	50	IUD
42	44	41	Dagang	PNS	D3	S1	6	60	IUD
43	42	36	Buruh	Swasta	SMP	SMA	3	33	Implant
44	27	20	Buruh	Swasta	SMP	SMP	1	6	Pil
45	32	28	Wiraswasta	PNS	SMP	SD	1	14	Pil
46	37	34	Petani	Swasta	SMA	D3	2	20	Suntik
47	27	22	Wiraswasta	Swasta	SMP	S1	1	7	Pil
48	29	23	Wiraswasta	Swasta	SMP	SMP	1	7	Pil
49	46	41	Petani	IRT	SMP	SMP	6	40	IUD
50	39	33	Wiraswasta	IRT	SMP	SMA	3	22	Suntik
51	29	26	PNS	IRT	SD	D3	1	12	Pil
52	29	21	Dagang	Swasta	SMA	SMA	0	0	Kondom
53	33	29	PNS	Dagang	SD	SMP	2	14	Pil
54	36	29	Dagang	Swasta	D3	SD	1	12	Pil
55	42	39	Swasta	IRT	SMA	SMA	4	27	Implant
56	47	44	Buruh	IRT	SMP	SMP	4	44	IUD
57	47	44	PNS	IRT	D3	SD	6	48	IUD
58	48	43	Swasta	PNS	S1	D3	6	55	IUD
59	47	42	Buruh	Dagang	S1	SMP	4	36	IUD
60	37	30	Buruh	Dagang	SMA	S1	1	11	Pil
61	41	34	Buruh	Dagang	SD	D3	2	24	Pil
62	52	44	Swasta	PNS	SMA	S1	4	50	IUD
63	28	23	Swasta	Swasta	SD	S1	1	9	Pil
64	37	30	Buruh	IRT	SD	S1	1	11	Pil
65	32	26	PNS	IRT	SD	SD	2	11	Pil
66	43	40	Buruh	PNS	SMA	SMP	3	35	Implant
67	33	25	Swasta	IRT	SMA	SMP	1	9	Pil
68	27	22	Wiraswasta	Dagang	SMA	SMP	1	6	Pil
69	42	39	Dagang	PNS	D3	SMP	4	33	Implant
70	51	44	Dagang	PNS	SMA	S1	5	55	IUD

ID	Usia Suami	Usia Istri	Pekerjaan Suami	Pekerjaan Istri	Pendidikan Suami	Pendidikan Istri	Jumlah Anak	Usia Anak Terakhir (Bulan)	Pilihan KB
71	33	29	Petani	Dagang	S1	SMP	1	12	Pil
72	26	22	Wiraswasta	PNS	SMP	S1	0	0	Kondom
73	53	45	Swasta	IRT	SD	D3	4	48	IUD
74	25	22	Petani	IRT	SMP	SMA	0	0	Kondom
75	48	43	PNS	Dagang	SMP	S1	6	55	IUD
76	42	39	Dagang	IRT	SMP	SMA	4	30	Implant
77	42	34	Dagang	PNS	S1	S1	3	16	Suntik
78	29	21	Dagang	Swasta	SMP	S1	1	9	Pil
79	44	39	Wiraswasta	Swasta	S1	S1	4	24	Suntik
80	35	31	Petani	PNS	SD	SD	3	24	Suntik
81	35	31	Wiraswasta	Dagang	SD	S1	3	14	Suntik
82	36	30	Petani	Swasta	SD	SMP	1	14	Pil
83	44	37	Swasta	Dagang	S1	D3	4	24	Implant
84	49	44	Petani	IRT	SMA	SD	4	44	IUD
85	42	38	Wiraswasta	IRT	S1	SMA	4	32	Implant
86	32	24	Petani	Dagang	SD	S1	1	7	Pil
87	43	36	Petani	Dagang	SMA	D3	4	24	Implant
88	34	31	Swasta	Swasta	D3	SD	3	18	Suntik
89	39	36	Petani	PNS	SD	SMA	3	35	Implant
90	39	31	Wiraswasta	IRT	SD	SD	2	24	Suntik
91	26	20	Petani	PNS	D3	SD	0	0	Kondom
92	43	39	Petani	PNS	SMP	SMP	4	30	Implant
93	28	20	Buruh	Dagang	SD	SD	0	0	Kondom
94	37	31	Petani	Dagang	S1	SMA	2	24	Suntik
95	50	45	Wiraswasta	Dagang	S1	SMP	6	48	IUD
96	32	28	Wiraswasta	Dagang	D3	SD	2	14	Pil
97	35	28	Dagang	PNS	SD	S1	1	10	Pil
98	31	27	PNS	IRT	SMP	S1	1	12	Pil
99	37	32	Swasta	Swasta	SMA	SMA	2	14	Suntik
100	40	34	Dagang	PNS	SMA	SMP	3	22	Suntik
101	42	39	Wiraswasta	PNS	S1	S1	3	32	Implant
102	31	23	Swasta	Dagang	D3	S1	1	8	Pil
103	45	42	Buruh	Swasta	SMP	D3	4	40	IUD
104	32	27	Buruh	Swasta	SD	SD	1	14	Pil
105	48	43	Swasta	Dagang	S1	SD	4	55	IUD
106	46	40	Wiraswasta	PNS	SMA	SMA	3	33	Implant
107	42	35	Dagang	Swasta	D3	D3	3	16	Suntik
108	42	36	Wiraswasta	Dagang	S1	D3	4	24	Implant

ID	Usia Suami	Usia Istri	Pekerjaan Suami	Pekerjaan Istri	Pendidikan Suami	Pendidikan Istri	Jumlah Anak	Usia Anak Terakhir (Bulan)	Pilihan KB
109	47	41	Petani	IRT	SMA	D3	4	48	IUD
110	33	30	PNS	Swasta	SMP	SD	2	10	Pil
111	46	38	Petani	Swasta	SMP	SMA	3	30	Implant
112	24	21	Petani	IRT	SMP	SMP	0	0	Kondom
113	47	44	Petani	PNS	SMA	D3	5	60	IUD
114	34	30	Dagang	IRT	D3	SD	1	10	Pil
115	32	26	Dagang	Dagang	SMA	SMP	2	12	Pil
116	53	45	Swasta	PNS	SD	D3	4	44	IUD
117	27	22	Buruh	IRT	D3	SMA	0	0	Kondom
118	36	31	Wiraswasta	Dagang	SMA	SMA	3	20	Suntik
119	35	27	Dagang	IRT	SD	D3	1	10	Pil
120	28	23	Swasta	Dagang	SMP	SMP	0	0	Kondom
121	42	37	Wiraswasta	PNS	SMP	SD	4	30	Implant
122	34	28	PNS	Swasta	SMP	S1	1	10	Pil
123	34	29	Swasta	Swasta	SD	SMP	2	12	Pil
124	32	27	Petani	IRT	SMP	SMA	2	10	Pil
125	29	22	Petani	IRT	S1	SMP	1	9	Pil
126	31	24	Petani	Swasta	D3	S1	1	8	Pil
127	32	26	Petani	Dagang	S1	SD	1	12	Pil
128	53	45	Petani	IRT	S1	SD	4	36	IUD
129	34	28	Buruh	IRT	SD	SMP	1	11	Pil
130	25	22	Buruh	IRT	SMA	SMA	1	7	Pil
131	42	39	Swasta	Swasta	D3	SMP	4	30	Implant
132	35	31	PNS	Dagang	SMA	S1	2	24	Suntik
133	29	26	Wiraswasta	Dagang	SMP	S1	2	10	Pil
134	43	37	PNS	Swasta	SD	SD	3	32	Implant
135	49	45	Petani	Swasta	SMP	D3	4	50	IUD

Data akseptor Keluarga Berencana (KB) dalam penelitian ini terdiri atas 135 observasi yang merepresentasikan variasi karakteristik pasangan usia subur (PUS). Setiap data memuat atribut demografis dan sosial, yaitu usia suami, usia istri, pekerjaan, tingkat pendidikan, jumlah anak, serta usia anak terakhir, dengan *Pilihan KB* sebagai label kelas.

Variasi usia mencerminkan perbedaan fase reproduksi dan pengalaman keluarga, sementara variabel pekerjaan dan pendidikan merepresentasikan kondisi sosial-ekonomi serta kapasitas akses informasi yang berpotensi memengaruhi keputusan kontrasepsi. Di sisi lain, jumlah anak dan usia anak terakhir menggambarkan dimensi pengalaman reproduksi yang berkaitan langsung dengan kebutuhan pengaturan kelahiran.

Dengan demikian, dataset ini tidak hanya berfungsi sebagai input komputasional, tetapi sebagai representasi empiris dari dinamika keputusan kontrasepsi dalam populasi yang beragam, yang selanjutnya dianalisis melalui pendekatan klasifikasi berbasis machine learning.

Seleksi Variabel dan Rasionalitas Representasi

Pemilihan variabel penelitian merupakan keputusan konseptual yang menentukan struktur ruang fitur dalam model klasifikasi. Variabel yang digunakan harus memiliki relevansi teoritis sekaligus representativitas empiris. Dalam konteks pemilihan alat kontrasepsi, variabel sosial-demografis dipilih karena secara konsisten diidentifikasi dalam literatur kesehatan reproduksi sebagai determinan keputusan.

Namun, seleksi variabel bukan sekadar proses administratif. Setiap variabel membawa implikasi interpretatif. Misalnya, usia dapat mencerminkan fase reproduksi, pengalaman keluarga, serta pertimbangan kesehatan. Pendidikan dapat merefleksikan literasi informasi dan akses pengetahuan. Jumlah anak dapat merepresentasikan kebutuhan pengaturan kelahiran. Dengan demikian, variabel dalam machine learning bukan hanya fitur matematis, tetapi konstruksi konseptual yang membentuk logika klasifikasi.

Seleksi data dilakukan dengan menentukan atribut/variabel yang relevan untuk membangun model klasifikasi pemilihan alat kontrasepsi. Variabel yang digunakan dalam penelitian ini dapat dilihat pada tabel berikut:

Tabel 2 Variabel yang Digunakan dalam Penelitian

No	Nama Variabel	Deskripsi	Tipe Data
1	Usia Suami	Umur suami dalam satuan tahun.	Numerik
2	Usia Istri	Umur istri dalam satuan tahun.	Numerik
3	Pekerjaan Suami	Jenis pekerjaan suami (misalnya Buruh, Petani, Dagang, Swasta, Wiraswasta, PNS).	Kategori
4	Pekerjaan Istri	Jenis pekerjaan istri (misalnya IRT, Swasta, Dagang, PNS).	Kategori
5	Pendidikan Suami	Tingkat pendidikan terakhir suami (SD, SMP, SMA, D3, S1).	Kategori
6	Pendidikan Istri	Tingkat pendidikan terakhir istri (SD, SMP, SMA, D3, S1).	Kategori

No	Nama Variabel	Deskripsi	Tipe Data
7	Jumlah Anak	Jumlah anak yang dimiliki pasangan.	Numerik
8	Usia Anak Terakhir (Bulan)	Umur anak terakhir dalam satuan bulan.	Numerik
9	Pilihan KB	Jenis alat kontrasepsi yang dipilih (Pil, Suntik, IUD, Implant, Kondom).	Kategori (Label)

Berdasarkan tabel 2, variabel yang digunakan terdiri dari atribut numerik dan atribut kategori. Atribut numerik seperti usia suami, usia istri, jumlah anak, dan usia anak terakhir (bulan) memberikan informasi kuantitatif mengenai kondisi keluarga. Sementara itu, atribut kategori seperti pekerjaan dan pendidikan suami/istri digunakan untuk menggambarkan kondisi sosial dan latar belakang pasangan secara kualitatif.

Seluruh variabel tersebut dipilih karena relevan dengan tujuan penelitian yaitu membangun sistem klasifikasi untuk memprediksi pemilihan alat kontrasepsi pada peserta KB. Data yang telah diseleksi kemudian diproses lebih lanjut pada tahap preprocessing agar siap digunakan dalam penerapan algoritma K-Nearest Neighbor (KNN).

Preprocessing Data sebagai Proses Epistemologis

Preprocessing sering dipahami sebagai prosedur teknis dalam pembersihan dan transformasi data. Namun dalam perspektif konseptual, preprocessing merupakan proses epistemologis yang menentukan bagaimana realitas direpresentasikan dalam model. Tahap ini mencakup pembersihan data, penanganan missing values, normalisasi, serta transformasi variabel kategorikal.

Cleaning data dilakukan untuk mengatasi permasalahan yang sering muncul pada data mentah seperti data kosong (missing value), data tidak konsisten, dan data duplikat. Proses cleaning dalam penelitian ini meliputi:

- Pemeriksaan Missing Value

Tahap ini dilakukan untuk memastikan setiap atribut seperti usia suami, usia istri, pekerjaan, pendidikan, jumlah anak, dan usia anak terakhir memiliki nilai yang lengkap. Data yang tidak lengkap dapat mempengaruhi hasil klasifikasi karena menyebabkan perhitungan jarak menjadi tidak akurat.

- **Pemeriksaan Konsistensi Data**

Pemeriksaan dilakukan untuk memastikan kesesuaian format dan logika data, misalnya:

- Usia suami dan usia istri harus dalam satuan tahun dan bernilai wajar.
- Jumlah anak tidak boleh bernilai negatif.
- Usia anak terakhir dalam bulan tidak boleh melebihi batas yang tidak masuk akal untuk kategori peserta KB.
- Penulisan kategori pekerjaan dan pendidikan harus konsisten, misalnya “Wiraswasta” tidak ditulis berbeda seperti “Wirausaha”.

- **Penghapusan Data Duplikat**

Data duplikat merupakan data yang memiliki isi atribut sama persis dengan baris lainnya. Data duplikat dapat menyebabkan model bias karena informasi yang sama dihitung berulang kali. Oleh karena itu, pada penelitian ini dilakukan pengecekan duplikasi berdasarkan seluruh atribut, dan hasilnya dataset yang digunakan sudah dipastikan tidak memiliki data duplikat.

Transformasi variabel kategorikal ke numerik juga membawa konsekuensi metodologis. Representasi angka terhadap kategori sosial tidak selalu memiliki makna jarak yang inheren. Oleh karena itu, keputusan transformasi perlu dipahami sebagai konstruksi analitik, bukan refleksi realitas sosial secara literal.

Tabel 3 Transformasi Variabel Pendidikan Suami/Istri

No	Variabel	Kategori	Kode
1	Pendidikan Suami/ Istri	SD	1
2	Pendidikan Suami/ Istri	SMP	2
3	Pendidikan Suami/ Istri	SMA	3
4	Pendidikan Suami/ Istri	D3	4
5	Pendidikan Suami/ Istri	S1	5

Tabel 4 Transformasi Variabel Pekerjaan Suami

No	Variabel	Kategori	Kode
1	Pekerjaan Suami	Buruh	1
2	Pekerjaan Suami	Petani	2
3	Pekerjaan Suami	Dagang	3
4	Pekerjaan Suami	Swasta	4
5	Pekerjaan Suami	Wiraswasta	5
6	Pekerjaan Suami	PNS	6

Tabel 5 Transformasi Variabel Pekerjaan Istri

No	Variabel	Kategori	Kode
1	Pekerjaan Istri	IRT	1
2	Pekerjaan Istri	Swasta	2
3	Pekerjaan Istri	Dagang	3
4	Pekerjaan Istri	PNS	4

Hasil dari proses Transformasi ini adalah dataset yang seluruh nilainya sudah berbentuk numerik sehingga dapat digunakan dalam proses perhitungan jarak pada algoritma KNN. Dataset hasil preprocessing kemudian digunakan pada tahap pembagian data training dan testing, serta dilanjutkan pada tahap klasifikasi pemilihan alat kontrasepsi.

Normalisasi dan Stabilitas Struktur Jarak

Dalam algoritma berbasis jarak seperti K-Nearest Neighbor, skala variabel memiliki pengaruh signifikan terhadap struktur kedekatan data. Variabel dengan rentang nilai besar berpotensi mendominasi perhitungan jarak, sehingga normalisasi menjadi prosedur penting untuk menjaga keseimbangan kontribusi fitur.

Normalisasi tidak sekadar prosedur matematis, tetapi mekanisme konseptual untuk memastikan bahwa setiap variabel diperlakukan secara proporsional dalam model. Tanpa normalisasi, interpretasi kedekatan dapat terdistorsi oleh dominasi variabel tertentu, yang berpotensi menurunkan validitas prediksi. Dalam konteks data sosial-demografis KB, normalisasi membantu menjaga integritas logika similarity antarindividu.

Pembagian Data Latih dan Data Uji sebagai Strategi Generalisasi

Pembagian dataset menjadi data latih (training) dan data uji (testing) merupakan strategi fundamental dalam machine learning. Tahap ini mencerminkan upaya model untuk belajar dari data historis sekaligus diuji terhadap data yang belum pernah diamati. Secara konseptual, proses ini merepresentasikan prinsip generalisasi, yaitu kemampuan model mengekstrapolasi pola ke situasi baru.

Rasio pembagian data memengaruhi stabilitas evaluasi. Dataset kecil memerlukan keseimbangan antara jumlah data pelatihan dan validitas pengujian. Pembagian data bukan sekadar prosedur teknis, tetapi refleksi dari asumsi mengenai representativitas pola populasi. Dalam penelitian ini, pembagian data memungkinkan evaluasi realistis terhadap performa model klasifikasi pemilihan kontrasepsi.

Jumlah data yang digunakan dalam penelitian ini sebanyak 135 data. Data tersebut kemudian dibagi menjadi data latih (training) dan data uji (testing) dengan perbandingan 80% : 20%, sehingga diperoleh 108 data latih dan 27 data uji. Pembagian data ini bertujuan untuk melatih model klasifikasi serta menguji kemampuan model dalam memprediksi data baru.

Tabel 4 Data Latih

ID	Usia Suami	Usia Istri	Pekerjaan Suami	Pekerjaan Istri	Pendidikan Suami	Pendidikan Istri	Jumlah Anak	Usia Anak Terakhir (Bulan)	Pilihan KB
2	39	32	6	1	5	4	2	20	Suntik
3	42	38	3	4	2	2	4	35	Implant
4	43	38	1	1	4	3	4	24	Implant
5	26	20	2	3	5	1	1	10	Pil
6	45	39	1	3	2	2	3	35	Implant
8	36	33	4	1	2	3	2	20	Suntik
9	42	36	2	1	2	4	4	33	Implant
11	52	44	2	4	1	1	4	40	IUD
13	28	23	2	1	5	3	0	0	Kondom
15	38	31	1	2	1	3	3	20	Suntik
16	44	40	6	3	1	3	3	35	Implant
17	33	27	3	4	1	4	2	12	Pil
19	41	33	5	2	3	4	3	16	Suntik
20	31	26	1	3	3	2	1	12	Pil
22	29	26	5	3	2	5	2	11	Pil
24	41	38	1	4	3	3	4	27	Implant
25	51	45	4	2	3	4	5	40	IUD
26	42	34	6	4	4	5	2	18	Suntik
27	34	28	5	4	2	4	2	11	Pil
29	46	43	5	3	1	5	5	48	IUD
30	31	24	1	4	1	1	1	10	Pil

31	49	44	2	4	1	4	4	60	IUD
33	39	31	5	1	1	2	2	16	Suntik
34	28	21	6	4	2	3	0	0	Kondom
35	51	44	2	3	1	3	6	48	IUD
36	33	25	1	2	4	5	0	0	Kondom
37	44	41	2	4	5	1	4	60	IUD
38	39	34	3	3	5	1	3	14	Suntik
39	29	25	1	3	5	3	1	6	Pil

ID	Usia Suami	Usia Istri	Pekerjaan Suami	Pekerjaan Istri	Pendidikan Suami	Pendidikan Istri	Jumlah Anak	Usia Anak Terakhir (Bulan)	Pilihan KB
40	48	45	1	3	2	5	6	44	IUD
41	46	41	1	4	3	5	4	50	IUD
42	44	41	3	4	4	5	6	60	IUD
45	32	28	5	4	2	1	1	14	Pil
46	37	34	2	2	3	4	2	20	Suntik
47	27	22	5	2	2	5	1	7	Pil
48	29	23	5	2	2	2	1	7	Pil
49	46	41	2	1	2	2	6	40	IUD
50	39	33	5	1	2	3	3	22	Suntik
51	29	26	6	1	1	4	1	12	Pil
52	29	21	3	2	3	3	0	0	Kondom
53	33	29	6	3	1	2	2	14	Pil
54	36	29	3	2	4	1	1	12	Pil
55	42	39	4	1	3	3	4	27	Implant
57	47	44	6	1	4	1	6	48	IUD
58	48	43	4	4	5	4	6	55	IUD
59	47	42	1	3	5	2	4	36	IUD
60	37	30	1	3	3	5	1	11	Pil
61	41	34	1	3	1	4	2	24	Pil
62	52	44	4	4	3	5	4	50	IUD
63	28	23	4	2	1	5	1	9	Pil
64	37	30	1	1	1	5	1	11	Pil
65	32	26	6	1	1	1	2	11	Pil
66	43	40	1	4	3	2	3	35	Implant
67	33	25	4	1	3	2	1	9	Pil
68	27	22	5	3	3	2	1	6	Pil
70	51	44	3	4	3	5	5	55	IUD
71	33	29	2	3	5	2	1	12	Pil
73	53	45	4	1	1	4	4	48	IUD
74	25	22	2	1	2	3	0	0	Kondom
75	48	43	6	3	2	5	6	55	IUD
76	42	39	3	1	2	3	4	30	Implant
78	29	21	3	2	2	5	1	9	Pil
79	44	39	5	2	5	5	4	24	Suntik
80	35	31	2	4	1	1	3	24	Suntik
81	35	31	5	3	1	5	3	14	Suntik
82	36	30	2	2	1	2	1	14	Pil
83	44	37	4	3	5	4	4	24	Implant
84	49	44	2	1	3	1	4	44	IUD
85	42	38	5	1	5	3	4	32	Implant
86	32	24	2	3	1	5	1	7	Pil
87	43	36	2	3	3	4	4	24	Implant

ID	Usia Suami	Usia Istri	Pekerjaan Suami	Pekerjaan Istri	Pendidikan Suami	Pendidikan Istri	Jumlah Anak	Usia Anak Terakhir (Bulan)	Pilihan KB
88	34	31	4	2	4	1	3	18	Suntik
89	39	36	2	4	1	3	3	35	Implant
91	26	20	2	4	4	1	0	0	Kondom
92	43	39	2	4	2	2	4	30	Implant
93	28	20	1	3	1	1	0	0	Kondom
94	37	31	2	3	5	3	2	24	Suntik
95	50	45	5	3	5	2	6	48	IUD
96	32	28	5	3	4	1	2	14	Pil
97	35	28	3	4	1	5	1	10	Pil
99	37	32	4	2	3	3	2	14	Suntik
101	42	39	5	4	5	5	3	32	Implant
102	31	23	4	3	4	5	1	8	Pil
103	45	42	1	2	2	4	4	40	IUD
104	32	27	1	2	1	1	1	14	Pil
106	46	40	5	4	3	3	3	33	Implant
107	42	35	3	2	4	4	3	16	Suntik
108	42	36	5	3	5	4	4	24	Implant
109	47	41	2	1	3	4	4	48	IUD
111	46	38	2	2	2	3	3	30	Implant
112	24	21	2	1	2	2	0	0	Kondom
113	47	44	2	4	3	4	5	60	IUD
114	34	30	3	1	4	1	1	10	Pil
115	32	26	3	3	3	2	2	12	Pil
116	53	45	4	4	1	4	4	44	IUD
117	27	22	1	1	4	3	0	0	Kondom
118	36	31	5	3	3	3	3	20	Suntik
119	35	27	3	1	1	4	1	10	Pil
120	28	23	4	3	2	2	0	0	Kondom
122	34	28	6	2	2	5	1	10	Pil
123	34	29	4	2	1	2	2	12	Pil
127	32	26	2	3	5	1	1	12	Pil
128	53	45	2	1	5	1	4	36	IUD
129	34	28	1	1	1	2	1	11	Pil
130	25	22	1	1	3	3	1	7	Pil
132	35	31	6	3	3	5	2	24	Suntik
133	29	26	5	3	2	5	2	10	Pil
135	49	45	2	2	2	4	4	50	IUD

Data latih yang digunakan dalam penelitian ini berjumlah 108 data yang diambil dari keseluruhan dataset pasangan usia subur (PUS) sebanyak 135 data. Data latih diperoleh dengan metode acak (random). Setiap data latih merepresentasikan karakteristik pasangan usia subur yang meliputi usia suami, usia istri, pekerjaan suami, pekerjaan istri, pendidikan suami, pendidikan istri, jumlah anak, serta usia anak terakhir dalam satuan bulan. Atribut pekerjaan dan pendidikan telah melalui proses

transformasi ke dalam bentuk numerik untuk memudahkan perhitungan jarak pada algoritma K-Nearest Neighbor (KNN), sedangkan variabel pilihan alat kontrasepsi tetap digunakan dalam bentuk kategorik sebagai label kelas. Data latih ini digunakan sebagai acuan utama dalam proses pembelajaran model KNN untuk mengenali pola hubungan antara karakteristik pasangan usia subur dengan jenis alat kontrasepsi yang dipilih, sehingga model mampu melakukan klasifikasi terhadap data uji secara lebih akurat.

Tabel 4 Data Uji

ID	Usia Suami	Usia Istri	Pekerjaan Suami	Pekerjaan Istri	Pendidikan Suami	Pendidikan Istri	Jumlah Anak	Usia Anak Terakhir (Bulan)
1	38	32	2	2	2	2	2	14
7	43	39	5	2	1	1	3	27
10	48	44	3	4	1	2	6	60
12	31	25	2	2	5	3	1	7
14	26	21	2	4	4	3	1	9
18	37	29	3	4	5	2	2	14
21	51	43	2	4	3	4	4	36
23	49	43	4	3	2	1	4	60
28	41	37	4	1	1	3	4	24
32	39	35	3	3	3	1	2	22
43	42	36	1	2	2	3	3	33
44	27	20	1	2	2	2	1	6
56	47	44	1	1	2	2	4	44
69	42	39	3	4	4	2	4	33
72	26	22	5	4	2	5	0	0
77	42	34	3	4	5	5	3	16
90	39	31	5	1	1	1	2	24
98	31	27	6	1	2	5	1	12
100	40	34	3	4	3	2	3	22
105	48	43	4	3	5	1	4	55
110	33	30	6	2	2	1	2	10
121	42	37	5	4	2	1	4	30
124	32	27	2	1	2	3	2	10
125	29	22	2	1	5	2	1	9
126	31	24	2	2	4	5	1	8
131	42	39	4	2	4	2	4	30
134	43	37	6	2	1	1	3	32

Data uji yang digunakan dalam penelitian ini berjumlah 27 data, yang diambil dari keseluruhan dataset pasangan usia subur (PUS) sebanyak 135 data. Data uji ditentukan secara acak (random). Data uji memiliki atribut yang sama dengan data latih, meliputi usia suami, usia istri, pekerjaan suami, pekerjaan istri, pendidikan suami, pendidikan istri, jumlah anak, serta usia anak terakhir dalam satuan bulan. Atribut pekerjaan dan pendidikan pada data uji telah ditransformasikan ke dalam bentuk numerik, sedangkan variabel pilihan alat kontrasepsi tetap digunakan dalam bentuk kategorik sebagai label kelas. Data uji ini digunakan untuk mengukur kinerja algoritma K-Nearest Neighbor (KNN) dalam mengklasifikasikan jenis alat kontrasepsi berdasarkan pola yang telah

dipelajari dari data latih, sehingga dapat diketahui tingkat akurasi dan keandalan model dalam melakukan prediksi pada data baru.

Risiko Distorsi dan Kesadaran Metodologis

Desain penelitian berbasis data sosial tidak terlepas dari risiko distorsi. Ketidakseimbangan kelas, korelasi variabel, serta keterbatasan ukuran dataset dapat memengaruhi performa model. Kesadaran metodologis diperlukan untuk menghindari overinterpretasi hasil klasifikasi.

Model machine learning bekerja berdasarkan pola statistik, bukan kausalitas sosial absolut. Oleh karena itu, interpretasi hasil harus mempertimbangkan bahwa prediksi merupakan estimasi kecenderungan berbasis data historis. Perspektif reflektif ini penting untuk menjaga integritas akademik monograf.

Desain penelitian dalam klasifikasi pemilihan alat kontrasepsi berfungsi sebagai kerangka rasionalitas analitik yang menentukan validitas model. Dataset akseptor KB merepresentasikan fenomena sosial-demografis yang kompleks dan memerlukan preprocessing sebagai proses epistemologis. Transformasi variabel dan normalisasi menentukan struktur jarak dalam algoritma KNN. Pembagian data latih dan uji mencerminkan prinsip generalisasi model. Kesadaran terhadap potensi distorsi menjadi prasyarat dalam interpretasi hasil analitik.

Variabel dalam machine learning harus dipahami sebagai konstruksi konseptual, bukan sekadar fitur numerik. Preprocessing merupakan proses pembentukan representasi realitas dalam model komputasional. Normalisasi menjaga stabilitas logika kedekatan data. Desain penelitian menentukan integritas inferensi prediktif. Kesadaran metodologis menjadi landasan penting dalam penggunaan model klasifikasi pada domain sosial-kesehatan.



BAB 7

IMPLEMENTASI MODEL KNN MENGGUNAKAN ORANGE DATA MINING

Implementasi Model sebagai Jembatan Konseptual dan Operasional

Implementasi model dalam machine learning merupakan fase di mana konstruksi konseptual diterjemahkan ke dalam prosedur komputasional. Tahap ini tidak sekadar proses teknis menjalankan algoritma, tetapi mekanisme validasi atas asumsi metodologis yang telah dibangun sebelumnya. Dalam penelitian klasifikasi pemilihan alat kontrasepsi pada Program Keluarga Berencana (KB), implementasi model berfungsi untuk menguji sejauh mana struktur data mampu menghasilkan pola prediktif yang stabil.

Penggunaan perangkat lunak analitik seperti Orange Data Mining memungkinkan eksplorasi model secara visual dan sistematis. Orange tidak hanya menyediakan antarmuka pemodelan, tetapi juga lingkungan eksperimental yang membantu peneliti memahami relasi antara data, algoritma, dan hasil prediksi. Dalam konteks akademik, implementasi model harus dipahami sebagai proses analitik yang mengintegrasikan logika algoritmik dan interpretasi konseptual.

Orange Data Mining dalam Ekosistem Analitik Modern

Orange Data Mining merepresentasikan pendekatan visual programming dalam machine learning. Lingkungan ini dirancang untuk memfasilitasi eksperimen analitik tanpa ketergantungan pada kode pemrograman kompleks. Namun, kemudahan

operasional tersebut tidak menghilangkan kebutuhan akan pemahaman metodologis yang mendalam.

Dalam perspektif epistemologis, Orange berfungsi sebagai medium representasi model. Setiap widget dalam workflow mencerminkan tahapan konseptual dalam machine learning, mulai dari input data, seleksi variabel, pelatihan model, hingga evaluasi performa. Dengan demikian, workflow bukan sekadar rangkaian prosedur teknis, melainkan struktur logika analitik yang merefleksikan desain penelitian.

Representasi Data dalam Lingkungan Komputasional

Tahap awal implementasi dimulai dengan pemuatan dataset hasil preprocessing. Dataset yang telah melalui proses cleaning, transformasi, dan normalisasi direpresentasikan dalam sistem sebagai struktur fitur numerik dan label kelas. Pada fase ini, integritas representasi data menjadi krusial, karena kesalahan format atau tipe variabel berpotensi menghasilkan distorsi model.

Pemahaman terhadap representasi data dalam lingkungan komputasional penting untuk menjaga validitas analisis. Data bukan sekadar input pasif, tetapi fondasi yang menentukan perilaku algoritma. Dalam K-Nearest Neighbor (KNN), struktur numerik data secara langsung memengaruhi perhitungan jarak dan mekanisme klasifikasi.

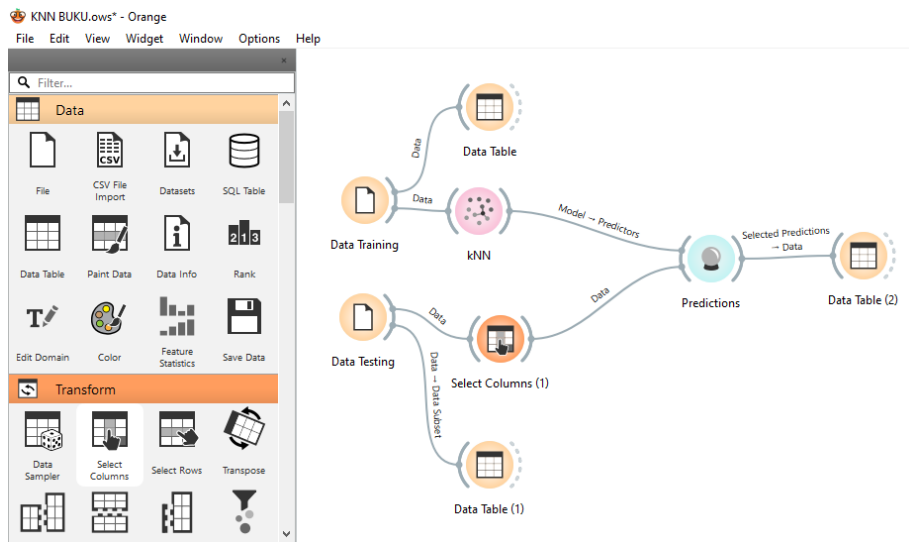
Pembentukan Model KNN: Logika Kedekatan dalam Praktik

Proses pembentukan model KNN dalam Orange merepresentasikan operasionalisasi prinsip similarity berbasis jarak. Algoritma bekerja dengan menghitung kedekatan antara data uji dan data latih dalam ruang fitur multidimensi. Tidak terdapat fase pelatihan parametrik kompleks; model terbentuk melalui struktur relasional antar data.

Proses pembagian dataset umumnya dilakukan dengan cara acak (random split), di mana data dibagi secara acak menjadi dua bagian. Sebagai aturan umum, 70%-80% dari data digunakan untuk training, dan sisanya 20%-30% digunakan untuk testing. Proporsi ini sering digunakan karena memberikan keseimbangan yang baik antara jumlah data yang digunakan untuk melatih model dan data yang cukup untuk melakukan pengujian[12].

Implementasi algoritma KNN dilakukan menggunakan aplikasi Orange Data Mining. Dataset hasil transformasi yang berjumlah 135 data dimasukkan ke dalam Orange melalui widget File. Data tersebut kemudian dibagi menjadi data latih dan data uji dengan perbandingan 80% untuk data training (108 data) dan 20% untuk data

testing (27 data). Alur pengolahan data pada Orange disusun dalam bentuk workflow yang saling terhubung antar-widget seperti berikut:



Gambar Rancangan Sistem

Input Data Training

Proses awal dimulai dengan memasukkan data training ke dalam sistem melalui widget Data pada aplikasi Orange. Data training merupakan kumpulan data yang digunakan untuk melatih model K-Nearest Neighbor (KNN) agar mampu mengenali pola klasifikasi berdasarkan atribut yang telah ditentukan. Dataset ini memuat variabel penelitian yang terdiri atas usia, tingkat pendidikan, usia anak terakhir (bulan), serta jenis alat kontrasepsi sebagai label kelas.

Data yang telah dimasukkan kemudian dihubungkan ke widget Data Table untuk menampilkan seluruh isi dataset dalam bentuk tabel. Tahap ini bertujuan untuk memastikan bahwa struktur data, tipe variabel, serta pembagian atribut dan label kelas telah terbaca dengan benar oleh sistem sebelum proses pelatihan model dilakukan. Verifikasi ini penting agar tidak terjadi kesalahan dalam proses klasifikasi pada tahap selanjutnya.

Pembentukan Model KNN

Dataset training selanjutnya dihubungkan ke widget KNN untuk membangun model klasifikasi. Pada tahap ini, algoritma K-Nearest Neighbor bekerja dengan cara menghitung jarak kedekatan antar data berdasarkan atribut yang dimiliki. Prinsip kerja KNN adalah menentukan nilai parameter K (jumlah tetangga terdekat), menghitung jarak data uji ke data latih, mengambil K data dengan jarak terdekat dan menentukan

kelas berdasarkan mayoritas tetangga. Model yang terbentuk dari proses ini kemudian disalurkan ke widget Predictions sebagai model prediktor.

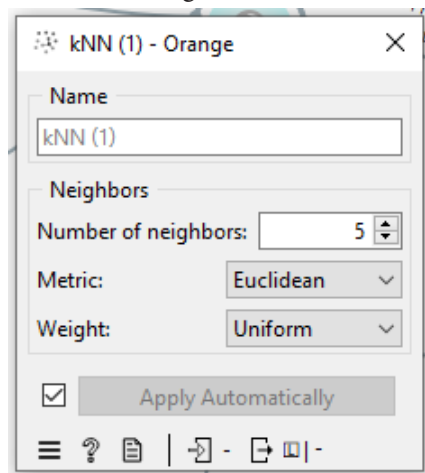
Input Data Testing

Selain data training, penelitian ini juga menggunakan data testing yang dimasukkan melalui widget Data Testing. Data testing merupakan data baru yang belum diketahui kelasnya dan digunakan untuk menguji kinerja model KNN. Data testing kemudian diproses pada widget Select Columns untuk menentukan atribut yang digunakan dalam proses pengujian. Pada tahap ini dilakukan pemilihan ; Feature (variabel input), Target/label (kelas) dan Meta (jika ada). Proses seleksi atribut bertujuan agar variabel yang digunakan sesuai dengan model training.

Nilai parameter K menjadi determinan penting dalam stabilitas model. Pemilihan K mencerminkan keputusan konseptual mengenai cakupan kedekatan yang dianggap relevan dalam menentukan kelas. Nilai K kecil meningkatkan sensitivitas lokal, sementara nilai K besar memperkuat generalisasi global. Oleh karena itu, implementasi model bukan sekadar menjalankan algoritma, tetapi refleksi atas asumsi metodologis yang mendasarinya.

Proses Prediksi

Setelah model KNN dan data testing siap, keduanya dihubungkan ke widget Predictions. Pada tahap ini sistem melakukan proses klasifikasi, yaitu memprediksi kelas data testing berdasarkan model yang telah dilatih sebelumnya. Output dari proses ini menampilkan kelas aktual, kelas hasil prediksi dan tingkat kecocokan.



Gambar Pembentukan Algoritma KNN

Nilai $K = 5$ dipilih pada algoritma K-Nearest Neighbor untuk memperoleh keseimbangan antara akurasi dan stabilitas model klasifikasi. Nilai K yang terlalu kecil cenderung sensitif terhadap noise, sedangkan nilai K yang terlalu besar dapat mengurangi kemampuan model dalam menangkap pola data. Penggunaan nilai $K = 5$ sebagai bilangan ganjil juga membantu menghindari hasil klasifikasi yang imbang, sehingga model dapat menghasilkan prediksi yang lebih konsisten. Berdasarkan hasil pengujian pada Orange, nilai $K = 5$ memberikan kinerja yang cukup baik sehingga digunakan dalam penelitian ini.

Prediksi dan Inferensi Berbasis Pola Historis

Tahap prediksi merupakan fase inferensi di mana model mengklasifikasikan data baru berdasarkan pola historis. Dalam konteks pemilihan alat kontrasepsi, prediksi tidak dimaksudkan untuk menentukan keputusan individu, melainkan menghasilkan estimasi berbasis kemiripan karakteristik sosial-demografis.

Interpretasi hasil prediksi memerlukan kehati-hatian konseptual. Model KNN bekerja berdasarkan kedekatan statistik, bukan kausalitas sosial. Oleh karena itu, hasil klasifikasi harus dipahami sebagai representasi kecenderungan, bukan determinasi absolut. Perspektif ini penting untuk menjaga integritas akademik dan etika penggunaan model prediktif dalam domain kesehatan reproduksi.

Hasil Prediksi

Hasil prediksi kemudian ditampilkan melalui widget Data Table (2). Tabel ini berisi data testing yang telah dilengkapi dengan hasil klasifikasi dari algoritma KNN. Melalui tampilan ini peneliti dapat melihat secara rinci data yang diprediksi, hasil kelas prediksi dan perbandingan dengan data asli. Selain itu, dataset hasil seleksi atribut juga ditampilkan pada Data Table (1) sebagai bagian dari proses validasi pemilihan variabel.

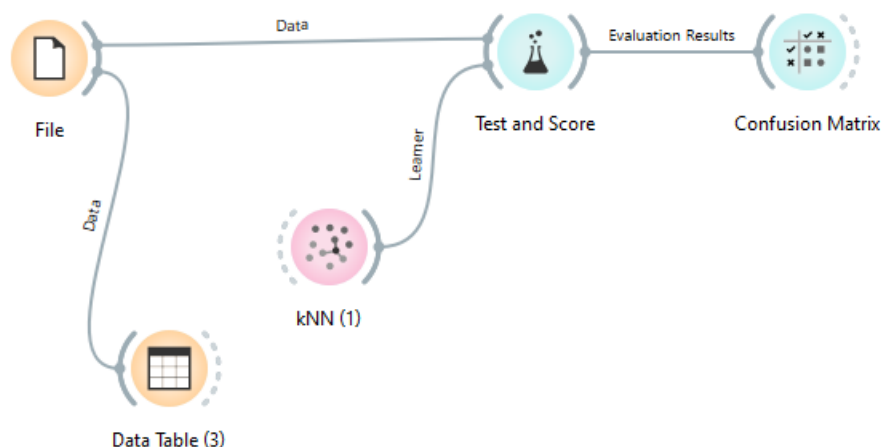
KNN	ID	Usia Suami	Usia Istri	Pekejaan Suami	Pekejaan Istri	Pendidikan Suami	Pendidikan Istri	Jumlah Anak	Anak Terakhir (B)
1	Suntik	77	42	34	3	4	5	5	3
2	Pil	12	31	25	2	2	5	3	1
3	Suntik	100	40	34	3	4	3	2	3
4	IUD	56	47	44	1	1	2	2	4
5	Suntik	1	38	32	2	2	2	2	2
6	Implant	121	42	37	5	4	2	1	4
7	Suntik	32	39	35	3	3	3	1	2
8	Suntik	90	39	31	5	1	1	1	2
9	Pil	125	29	22	2	1	5	2	1
10	Pil	126	31	24	2	2	4	5	1
11	Pil	124	32	27	2	1	2	3	2
12	Pil	18	37	29	2	4	5	2	4
13	IUD	21	51	43	2	4	3	4	4
14	Implant	7	43	39	5	2	1	1	3
15	Pil	110	33	30	6	2	2	1	2
16	Pil	44	27	20	1	2	2	2	1
17	IUD	43	42	36	1	2	2	3	3
18	Pil	72	26	22	5	4	2	5	0
19	IUD	131	42	39	4	2	4	2	4
20	Pil	98	31	27	6	1	2	5	1
21	IUD	10	48	44	3	4	1	2	6
22	Suntik	28	41	37	4	1	1	3	4
23	IUD	134	43	37	6	2	1	1	3
24	IUD	105	48	43	4	3	5	1	4
25	IUD	23	49	43	4	3	2	1	4
26	Implant	69	42	39	3	4	4	2	4
27	Pil	14	26	21	2	4	4	3	1

Gambar 4.3 Hasil Prediksi

Gambar tersebut menunjukkan hasil prediksi klasifikasi data uji menggunakan algoritma K-Nearest Neighbor (KNN) pada aplikasi Orange. Setiap data uji diklasifikasikan berdasarkan kedekatan jarak terhadap data latih menggunakan atribut usia suami, usia istri, pekerjaan suami, pekerjaan istri, pendidikan suami, pendidikan istri, jumlah anak, dan usia anak terakhir. Kolom kNN merepresentasikan kelas hasil prediksi yang diperoleh dari mayoritas tetangga terdekat. Hasil prediksi ini selanjutnya digunakan untuk membangun Confusion Matrix dan menghitung nilai akurasi, precision, recall, serta F1-score guna mengevaluasi performa model.

Evaluasi Model sebagai Proses Validasi Epistemologis

Evaluasi model dalam Orange dilakukan melalui metrik performa seperti akurasi, precision, recall, dan F1-score. Namun, evaluasi bukan sekadar pengukuran numerik, melainkan proses validasi epistemologis terhadap kemampuan model merepresentasikan pola data.



Confusion matrix memberikan perspektif yang lebih kaya dibanding akurasi tunggal. Matriks ini memungkinkan identifikasi distribusi kesalahan klasifikasi antar kelas. Dalam penelitian KB, analisis confusion matrix membantu memahami dinamika prediksi, misalnya kecenderungan model mengklasifikasikan metode tertentu secara lebih akurat dibanding metode lain. Penggunaan beberapa metrik evaluasi mencerminkan kesadaran bahwa performa model bersifat multidimensional.

Gambar 4.4 Rancangan Sistem *Confusion Matrix*

Gambar tersebut menunjukkan alur proses klasifikasi pemilihan alat kontrasepsi menggunakan algoritma *K-Nearest Neighbor* (KNN) pada aplikasi Orange. Proses

diawali dengan pemuatan dataset melalui widget *File*, yang berfungsi untuk mengimpor data penelitian ke dalam sistem. Dataset yang telah diimpor kemudian dapat ditampilkan menggunakan widget *Data Table* untuk memastikan bahwa data telah terbaca dengan benar sebelum dilakukan proses analisis lebih lanjut.

Selanjutnya, data dari widget *File* dihubungkan ke widget *KNN* sebagai learner untuk membangun model klasifikasi. Model KNN yang telah terbentuk kemudian dihubungkan ke widget *Test and Score* untuk dilakukan pengujian dan evaluasi kinerja model. Pada tahap ini, sistem menghitung berbagai metrik evaluasi seperti akurasi, precision, recall, dan F1-score guna menilai kemampuan model dalam melakukan klasifikasi. Hasil evaluasi tersebut kemudian ditampilkan dalam bentuk *Confusion Matrix*, yang digunakan untuk mengetahui jumlah prediksi yang benar dan salah pada setiap kelas. Alur ini menunjukkan bahwa proses klasifikasi dilakukan secara sistematis mulai dari pemuatan data, pembentukan model, hingga evaluasi hasil klasifikasi.

Sensitivitas Model dan Refleksi Metodologis

Implementasi model KNN memperlihatkan bahwa performa algoritma sangat dipengaruhi oleh struktur dataset. Distribusi kelas, skala variabel, serta representasi numerik variabel kategorikal membentuk perilaku model secara signifikan. Sensitivitas ini menegaskan bahwa machine learning bukan proses mekanistik, melainkan sistem analitik yang bergantung pada kualitas representasi data.

Refleksi metodologis diperlukan untuk menghindari interpretasi yang simplistik. Model dengan akurasi moderat tetap memiliki nilai analitik jika mampu mengungkap pola yang relevan secara konseptual. Sebaliknya, akurasi tinggi tidak selalu menjamin validitas substantif jika diperoleh dari distorsi data.

Implementasi model KNN menggunakan Orange Data Mining merepresentasikan penerjemahan konsep similarity ke dalam prosedur komputasional. Workflow analitik mencerminkan struktur logika machine learning yang sistematis. Representasi data menentukan perilaku algoritma, sementara parameter K memengaruhi stabilitas model. Prediksi berbasis KNN menghasilkan estimasi kecenderungan, bukan determinasi kausal. Evaluasi model melalui berbagai metrik memberikan gambaran performa yang komprehensif. Sensitivitas model menegaskan pentingnya refleksi metodologis dalam interpretasi hasil.

Implementasi model merupakan proses epistemologis yang memvalidasi struktur representasi data. Orange Data Mining berfungsi sebagai medium analitik visual yang memperjelas logika machine learning. Parameter model mencerminkan asumsi konseptual, bukan sekadar konfigurasi teknis. Evaluasi performa harus dipahami secara

multidimensional. Model prediktif berperan sebagai sistem pendukung keputusan, bukan pengganti pertimbangan profesional dalam layanan KB.



BAB 8

EVALUASI MODEL DAN INTERPRETASI KINERJA KLASIFIKASI

Evaluasi Model sebagai Proses Epistemologis

Dalam machine learning, evaluasi model sering dipersepsikan sebagai tahap akhir yang bersifat teknis, yaitu menghitung akurasi dan metrik performa lainnya. Namun dalam perspektif ilmiah yang lebih reflektif, evaluasi model merupakan proses epistemologis yang menentukan validitas pengetahuan yang dihasilkan sistem. Model klasifikasi pada dasarnya adalah konstruksi inferensial; kualitasnya tidak hanya ditentukan oleh ketepatan numerik, tetapi oleh relevansi interpretatif terhadap fenomena yang direpresentasikan.

Pada penelitian klasifikasi pemilihan alat kontrasepsi dalam Program Keluarga Berencana (KB), evaluasi model memiliki makna yang lebih luas dibanding sekadar mengukur performa algoritma. Evaluasi berfungsi untuk menilai sejauh mana struktur data sosial-demografis mampu menjelaskan kecenderungan pilihan kontrasepsi secara sistematis. Dengan demikian, metrik evaluasi tidak hanya mencerminkan kinerja komputasional, tetapi juga kapasitas representasi model terhadap realitas sosial.

Model	AUC	CA	F1	Prec	Recall	MCC
kNN (1)	0.990	0.963	0.963	0.964	0.963	0.952

Gambar Evaluasi *Test and Score*

Berdasarkan hasil evaluasi menggunakan *Test and Score* pada aplikasi Orange, model *K-Nearest Neighbor* (KNN) menghasilkan nilai *AUC* sebesar 0,894 atau 89%, yang menunjukkan bahwa model memiliki kemampuan yang sangat baik dalam membedakan antar kelas. Nilai *Classification Accuracy* (CA) sebesar 0,681 menunjukkan bahwa model mampu mengklasifikasikan data dengan tingkat ketepatan sebesar 68,1%. Nilai *F1-score* sebesar 0,672 menandakan keseimbangan yang baik antara nilai *precision* dan *recall* dalam proses klasifikasi.

Selain itu, nilai *Precision* sebesar 0,667 menunjukkan bahwa sebagian besar hasil prediksi yang dihasilkan oleh model adalah benar, sedangkan nilai *Recall* sebesar 0,681 menunjukkan bahwa model cukup baik dalam mengenali data pada kelas yang sesuai. Nilai *Matthews Correlation Coefficient* (MCC) sebesar 0,567 mengindikasikan adanya korelasi yang cukup kuat antara hasil prediksi model dengan data aktual. Secara keseluruhan, hasil ini menunjukkan bahwa algoritma KNN memiliki kinerja yang cukup baik dan layak digunakan dalam klasifikasi pemilihan alat kontrasepsi.

Keterbatasan Akurasi sebagai Indikator Tunggal

Akurasi klasifikasi (*Classification Accuracy / CA*) pada model KNN menunjukkan nilai **0,963**, yang mengindikasikan bahwa model mampu mengklasifikasikan 96,3% data secara benar. Nilai ini mencerminkan performa model yang sangat tinggi dan menunjukkan bahwa pola hubungan antar variabel dalam dataset relatif stabil serta dapat dikenali dengan baik oleh algoritma.

Meskipun demikian, akurasi tetap tidak dapat dipandang sebagai indikator tunggal dalam evaluasi model klasifikasi. Pada dataset multikelas seperti pemilihan alat kontrasepsi, akurasi yang tinggi belum tentu menggambarkan distribusi performa yang merata pada seluruh kategori. Model dapat menunjukkan performa global yang sangat baik, tetapi tetap berpotensi memiliki variasi sensitivitas pada kelas tertentu. Oleh karena itu, interpretasi kinerja model memerlukan metrik tambahan untuk memperoleh gambaran evaluatif yang lebih komprehensif.

Precision dan Makna Ketepatan Prediksi

Nilai *precision* sebesar **0,964** menunjukkan tingkat ketepatan prediksi yang sangat baik. *Precision* yang mendekati nilai maksimum mengindikasikan bahwa prediksi model terhadap kategori kontrasepsi memiliki tingkat kesesuaian yang sangat tinggi dengan label aktual. Artinya, ketika model menetapkan suatu metode sebagai hasil klasifikasi, kemungkinan prediksi tersebut benar sangat besar.

Dalam konteks analisis pemilihan kontrasepsi, precision yang tinggi merefleksikan reliabilitas model dalam meminimalkan kesalahan klasifikasi positif. Hal ini memperlihatkan bahwa model KNN mampu menjaga konsistensi prediksi meskipun karakteristik sosial-demografis antar pengguna metode dapat menunjukkan kemiripan tertentu.

Recall dan Sensitivitas Model terhadap Variasi Kelas

Nilai precision sebesar **0,964** menunjukkan tingkat ketepatan prediksi yang sangat baik. Precision yang mendekati nilai maksimum mengindikasikan bahwa prediksi model terhadap kategori kontrasepsi memiliki tingkat kesesuaian yang sangat tinggi dengan label aktual. Artinya, ketika model menetapkan suatu metode sebagai hasil klasifikasi, kemungkinan prediksi tersebut benar sangat besar.

Dalam konteks analisis pemilihan kontrasepsi, precision yang tinggi merefleksikan reliabilitas model dalam meminimalkan kesalahan klasifikasi positif. Hal ini memperlihatkan bahwa model KNN mampu menjaga konsistensi prediksi meskipun karakteristik sosial-demografis antar pengguna metode dapat menunjukkan kemiripan tertentu.

F1-Score sebagai Keseimbangan Evaluatif

Nilai F1-score sebesar **0,963** mencerminkan keseimbangan yang sangat baik antara precision dan recall. Nilai ini menunjukkan bahwa model memiliki stabilitas klasifikasi yang tinggi tanpa kecenderungan bias terhadap salah satu dimensi evaluasi. Dalam perspektif machine learning, F1-score yang tinggi mengindikasikan bahwa model mampu mempertahankan konsistensi performa pada berbagai kategori.

Nilai F1-score yang mendekati maksimum memperlihatkan bahwa pola dalam dataset dapat direpresentasikan secara efektif oleh algoritma KNN. Namun demikian, performa tinggi ini tetap perlu dipahami dalam konteks representasi data historis, di mana model bekerja berdasarkan struktur variabel yang tersedia.

Confusion Matrix dan Struktur Kesalahan Prediksi

Confusion matrix merupakan instrumen evaluasi yang memberikan gambaran paling kaya mengenai performa model klasifikasi. Matriks ini menampilkan distribusi prediksi benar dan salah pada setiap kelas. Diagonal utama menunjukkan prediksi yang tepat, sementara elemen di luar diagonal merepresentasikan kesalahan klasifikasi.

Dalam analisis pemilihan alat kontrasepsi, confusion matrix memungkinkan identifikasi pola kesalahan sistematis. Misalnya, jika model sering mengklasifikasikan

pengguna Implant sebagai Suntik, hal tersebut dapat mengindikasikan kedekatan karakteristik sosial-demografis antara kedua kelompok. Interpretasi semacam ini memperluas evaluasi dari sekadar angka menuju pemahaman pola. Dengan demikian, confusion matrix berfungsi sebagai alat interpretasi fenomenologis dalam machine learning.

		Predicted					Σ
		IUD	Implant	Kondom	Pil	Suntik	
Actual	IUD	30	0	0	0	0	30
	Implant	0	24	0	0	1	25
	Kondom	0	0	11	0	0	11
	Pil	0	0	0	44	2	46
	Suntik	0	1	0	1	21	23
Σ		30	25	11	45	24	135

Gambar 4.6 Evaluasi *Confusion Matrix*

Confusion Matrix pada gambar menunjukkan hasil klasifikasi pemilihan alat kontrasepsi menggunakan algoritma *K-Nearest Neighbor (KNN)* terhadap 135 data. *Confusion matrix* menunjukkan bahwa model *K-Nearest Neighbor (KNN)* memiliki kinerja klasifikasi yang sangat baik dengan dominasi prediksi benar pada diagonal utama. Kategori **IUD** memperlihatkan performa sempurna, di mana seluruh 30 observasi aktual berhasil diklasifikasikan secara tepat tanpa kesalahan. Pola serupa juga terlihat pada kategori **Kondom**, dengan 11 data aktual seluruhnya diprediksi dengan benar. Pada kategori **Implant**, dari 25 observasi aktual, sebanyak 24 diklasifikasikan dengan benar, sementara satu observasi mengalami kesalahan prediksi ke kategori **Suntik**, mengindikasikan adanya kedekatan karakteristik antara kedua metode tersebut dalam ruang fitur.

Kategori **Pil** menunjukkan tingkat akurasi yang sangat tinggi dengan 44 prediksi benar dari total 46 observasi aktual. Dua kesalahan prediksi pada kelas ini terjadi menuju kategori **Suntik**, yang kembali mengisyaratkan potensi kemiripan atribut sosial-demografis antara pengguna Pil dan Suntik. Pada kategori **Suntik**, model mengklasifikasikan 21 dari 23 observasi secara tepat, dengan dua kesalahan prediksi yang tersebar ke kategori **Implant** dan **Pil**. Distribusi kesalahan yang relatif kecil dan tidak terkonsentrasi pada satu kelas tertentu menunjukkan bahwa model tidak mengalami bias signifikan terhadap kategori dominan.

Secara keseluruhan, struktur confusion matrix mengindikasikan bahwa model memiliki stabilitas prediksi yang kuat, dengan kesalahan klasifikasi yang terbatas dan cenderung terjadi antar metode yang secara konseptual memiliki karakteristik berdekatan. Fenomena ini mencerminkan bahwa model KNN bekerja konsisten dalam mengenali pola kedekatan data, sekaligus menegaskan bahwa beberapa metode kontrasepsi memiliki profil atribut yang secara empiris sulit dipisahkan secara tegas.

AUC dan Dimensi Diskriminatif Model

Area Under Curve (AUC) mengukur kemampuan model dalam membedakan antar kelas. Nilai AUC yang tinggi menunjukkan kapasitas diskriminatif model yang baik. Dalam konteks multikelas, AUC merepresentasikan sejauh mana model mampu memisahkan kategori berdasarkan struktur fitur.

Nilai AUC sebesar **0,990** pada model KNN mengindikasikan kemampuan diskriminatif yang sangat tinggi, mendekati kondisi ideal. Angka ini menunjukkan bahwa model hampir selalu mampu memberikan separasi yang tepat antara kategori pemilihan alat kontrasepsi. Dengan kata lain, struktur variabel dalam dataset membentuk ruang fitur yang memungkinkan pemisahan pola berlangsung secara sangat jelas.

Dalam konteks klasifikasi multikelas pada penelitian Program Keluarga Berencana (KB), AUC yang sangat tinggi mencerminkan bahwa variabel sosial-demografis yang digunakan memiliki relevansi informasional yang kuat. Model tidak hanya akurat dalam prediksi, tetapi juga efektif dalam membedakan karakteristik antar pengguna metode kontrasepsi yang berbeda.

Lebih jauh, AUC tidak semata-mata berkaitan dengan ketepatan hasil klasifikasi, tetapi dengan kualitas pemetaan data dalam ruang fitur multidimensi. Nilai AUC yang mendekati maksimum menunjukkan bahwa representasi variabel penelitian mampu menciptakan struktur separasi yang stabil. Hal ini menegaskan bahwa pola dalam dataset relatif konsisten dan dapat dikenali secara sistematis oleh algoritma KNN.

Namun demikian, interpretasi nilai AUC tetap memerlukan perspektif reflektif. Performa diskriminatif yang sangat tinggi menunjukkan kesesuaian model terhadap dataset yang dianalisis, tetapi tidak serta-merta menjamin bahwa seluruh kompleksitas keputusan kontrasepsi di dunia nyata telah terwakili secara penuh dalam variabel yang digunakan.

Matthews Correlation Coefficient dan Stabilitas Prediksi

Matthews Correlation Coefficient (MCC) merupakan metrik evaluasi yang memperhitungkan keseimbangan distribusi kelas. MCC sering dipandang lebih robust dibanding akurasi karena sensitif terhadap ketidakseimbangan data. Nilai MCC yang positif menunjukkan korelasi yang bermakna antara prediksi model dan data aktual (Chicco & Jurman, 2020). Dalam penelitian sosial-demografis KB, MCC memberikan indikasi stabilitas model dalam menangani variasi kelas.

Nilai MCC sebesar **0,952** pada model KNN menunjukkan korelasi yang sangat kuat antara hasil prediksi dan label aktual. Nilai yang mendekati +1 mengindikasikan bahwa model tidak hanya menghasilkan akurasi tinggi secara agregat, tetapi juga mempertahankan konsistensi prediksi pada berbagai kelas kontrasepsi. Dengan kata lain, performa model relatif stabil dan tidak didominasi oleh kelas tertentu.

Dalam konteks penelitian sosial-demografis pada Program Keluarga Berencana (KB), nilai MCC yang sangat tinggi merefleksikan kemampuan model dalam menangani variasi karakteristik pengguna serta distribusi metode kontrasepsi. Hal ini memperlihatkan bahwa struktur fitur dalam dataset memiliki relevansi yang kuat terhadap proses klasifikasi, sehingga model mampu menjaga keseimbangan prediksi lintas kategori.

Namun demikian, interpretasi MCC tetap perlu ditempatkan dalam kerangka reflektif. Nilai yang sangat tinggi mencerminkan kesesuaian model terhadap dataset yang digunakan, bukan jaminan bahwa model sepenuhnya bebas dari keterbatasan representasi variabel atau dinamika konteks sosial di luar data.

Interpretasi Evaluasi: Antara Statistik dan Realitas Sosial

Interpretasi hasil evaluasi model memerlukan keseimbangan antara pemahaman statistik dan refleksi konseptual. Model machine learning bekerja pada representasi numerik realitas; kesalahan prediksi tidak selalu mencerminkan kegagalan algoritma, tetapi dapat merefleksikan kompleksitas fenomena sosial yang tidak sepenuhnya terukur.

Variabel sosial-demografis tidak selalu menangkap dimensi psikologis, budaya, dan relasional yang memengaruhi keputusan kontrasepsi. Oleh karena itu, evaluasi model harus dipahami sebagai evaluasi representasi data, bukan evaluasi realitas sosial secara absolut.

Evaluasi model klasifikasi merupakan proses epistemologis yang menentukan validitas pengetahuan prediktif. Akurasi sebagai indikator tunggal memiliki keterbatasan

pada dataset multikelas dan tidak seimbang. Precision, recall, dan F1-score memberikan perspektif evaluatif yang lebih komprehensif. Confusion matrix memungkinkan analisis pola kesalahan prediksi. AUC dan MCC merepresentasikan dimensi diskriminatif dan stabilitas model. Interpretasi evaluasi harus mempertimbangkan keterbatasan representasi data sosial.

Metrik evaluasi bukan sekadar angka performa, tetapi instrumen interpretasi struktur data. Kesalahan klasifikasi dapat merefleksikan kedekatan karakteristik sosial, bukan kegagalan model semata. Evaluasi model machine learning pada domain sosial-kesehatan menuntut perspektif reflektif. Validitas prediksi berkaitan erat dengan kualitas representasi variabel dan distribusi data.



BAB 9

REFLEKSI, SINTESIS, DAN ARAH PENGEMBANGAN ANALITIK PROGRAM KELUARGA BERENCANA

Klasifikasi Kontrasepsi sebagai Representasi Realitas Sosial

Klasifikasi pemilihan alat kontrasepsi melalui pendekatan machine learning memperlihatkan bahwa data dapat berfungsi sebagai medium untuk membaca keteraturan perilaku sosial. Namun demikian, penting untuk menegaskan bahwa model klasifikasi tidak merepresentasikan realitas sosial secara absolut. Model bekerja pada representasi numerik dari fenomena, bukan fenomena itu sendiri. Variabel sosial-demografis yang digunakan dalam analisis merupakan abstraksi terukur dari dinamika kehidupan yang jauh lebih kompleks.

Keputusan kontrasepsi pada pasangan usia subur (PUS) dipengaruhi oleh interaksi multidimensional antara kondisi ekonomi, persepsi risiko, pengalaman reproduksi, norma sosial, hingga preferensi personal. Dalam konteks ini, model klasifikasi berbasis K-Nearest Neighbor (KNN) berfungsi sebagai instrumen untuk mengenali pola kedekatan data, bukan mekanisme yang menentukan kebenaran tunggal. Klasifikasi merupakan pembacaan probabilistik terhadap struktur data historis yang mencerminkan kecenderungan, bukan determinasi perilaku individu. Perspektif ini penting untuk menjaga keseimbangan antara optimisme terhadap teknologi analitik dan kesadaran terhadap kompleksitas realitas sosial-kesehatan.

Kontribusi Metodologis Pendekatan Machine Learning

Pendekatan machine learning dalam analisis pemilihan alat kontrasepsi memberikan kontribusi metodologis yang signifikan terhadap studi Program Keluarga Berencana (KB). Secara konseptual, machine learning menghadirkan paradigma analitik yang berorientasi pada pola, bukan sekadar hubungan linear antar variabel. Algoritma KNN, dengan logika similarity berbasis jarak, memungkinkan identifikasi kecenderungan klasifikasi tanpa memerlukan asumsi distribusi parametrik yang kaku.

Kontribusi utama pendekatan ini terletak pada kemampuannya mengurangi subjektivitas dalam interpretasi data sosial-demografis. Model prediktif menyediakan kerangka inferensial yang konsisten dalam membaca konfigurasi atribut pengguna layanan. Dalam konteks pelayanan publik, konsistensi ini memiliki nilai strategis karena mendukung rasionalitas keputusan berbasis data.

Lebih jauh, penggunaan machine learning dalam domain sosial-kesehatan memperlihatkan bahwa teknik komputasional tidak terbatas pada analisis numerik, tetapi dapat berfungsi sebagai alat konseptual dalam memahami perilaku manusia melalui representasi data.

Batasan Epistemologis dan Risiko Interpretasi

Meskipun model klasifikasi menunjukkan performa evaluasi yang sangat tinggi, interpretasi hasil tetap memerlukan kehati-hatian epistemologis. Machine learning tidak menghasilkan pengetahuan kausal, melainkan inferensi berbasis pola statistik. Nilai akurasi, AUC, maupun MCC mencerminkan kesesuaian model terhadap dataset, bukan validitas universal terhadap seluruh konteks sosial.

Variabel yang digunakan dalam penelitian merupakan representasi parsial dari faktor-faktor yang memengaruhi keputusan kontrasepsi. Dimensi psikologis, budaya, relasional, serta pengalaman subjektif pengguna tidak sepenuhnya tertangkap dalam struktur data kuantitatif. Oleh karena itu, performa model yang tinggi tidak boleh diinterpretasikan sebagai reduksi kompleksitas keputusan reproduksi manusia ke dalam parameter algoritmik semata.

Kesadaran terhadap batas epistemologis model memperkuat integritas ilmiah sekaligus mencegah determinisme teknologi dalam analisis sosial.

Implikasi bagi Kebijakan dan Pelayanan Program KB

Integrasi analitik berbasis machine learning dalam Program KB membuka peluang penguatan sistem pendukung keputusan (decision support system). Model klasifikasi

berpotensi membantu tenaga kesehatan dan petugas lapangan dalam memahami kecenderungan pola pemilihan kontrasepsi berdasarkan karakteristik pengguna. Pemanfaatan ini dapat meningkatkan konsistensi evaluasi profil pengguna sekaligus memperkaya proses konseling.

Namun demikian, penggunaan model prediktif dalam layanan kesehatan reproduksi harus ditempatkan dalam kerangka etika dan profesionalisme. Sistem analitik berfungsi sebagai instrumen pendukung, bukan pengganti pertimbangan klinis dan komunikasi interpersonal. Keputusan akhir tetap berada pada interaksi antara pengguna layanan dan tenaga profesional kesehatan.

Pendekatan berbasis data, jika dikelola secara reflektif, dapat memperkuat rasionalitas kebijakan tanpa mengabaikan dimensi humanistik pelayanan publik.

Arah Pengembangan Analitik dan Penelitian Masa Depan

Masa depan analisis data dalam Program KB menuntut pengayaan baik pada dimensi data maupun metodologi. Pengembangan dataset dengan cakupan observasi yang lebih luas, representasi wilayah yang beragam, serta variabel perilaku dan persepsi akan memperkaya kapasitas model prediktif. Integrasi variabel laten seperti literasi kesehatan, preferensi individu, dan dinamika relasional berpotensi meningkatkan kedalaman interpretasi analitik.

Pada sisi metodologis, eksplorasi algoritma alternatif, model hibrida, serta pendekatan ensemble dapat menjadi arah pengembangan yang relevan. Perkembangan analitik tidak hanya bertujuan meningkatkan performa numerik, tetapi memperluas pemahaman konseptual terhadap struktur keputusan reproduksi dalam masyarakat.

Dengan demikian, evolusi analitik KB harus dipahami sebagai proses berkelanjutan yang mengintegrasikan data, teknologi, dan refleksi sosial.

Penutup Reflektif

Pemanfaatan machine learning dalam analisis pemilihan alat kontrasepsi merepresentasikan transformasi penting dalam cara memahami fenomena sosial-kesehatan melalui data. Model klasifikasi bukan sekadar alat teknis, melainkan medium konseptual yang memungkinkan pembacaan pola keputusan secara sistematis. Teknologi analitik tidak menggantikan kompleksitas realitas manusia, tetapi menyediakan perspektif baru dalam mengelola ketidakpastian informasi.

Pada akhirnya, nilai strategis pendekatan berbasis data terletak bukan pada klaim presisi absolut, melainkan pada kontribusinya dalam memperkaya rasionalitas,

konsistensi, dan refleksi kritis dalam kebijakan serta layanan Program Keluarga Berencana.

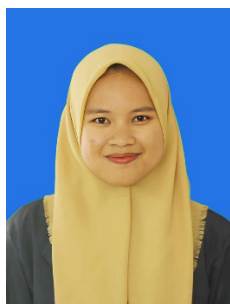
DAFTAR PUSTAKA

- Badan Pusat Statistik. (2023). Statistik Indonesia 2023. Badan Pusat Statistik. <https://www.bps.go.id>
- Badan Kependudukan dan Keluarga Berencana Nasional. (2022). Laporan Kinerja BKKBN 2022. BKKBN.
- World Health Organization. (2021). Global strategy on digital health 2020–2025. World Health Organization.
- N. Van Eekert et al., “Relationship between classic indicators of health behaviour and contraceptive choices in women in Flanders,” *BMC Womens. Health*, vol. 24, 2024, doi: 10.1186/s12905-024-03079-y.
- J. Bishop, B. Rosen, L. Rojas-Guyler, A. Bernard, and B. Wilson, “Intention to use novel methods of male directed contraception among college students,” *J. Sex. Heal. Psychol.*, vol. 1, no. 1, pp. 11–29, 2022, doi: 10.52547/jshp.1.1.11.
- E. al Reena Thakur, “A comprehensive analysis to image classification: Understanding techniques and explore data preprocessing a non-linear approach,” *Adv. Nonlinear Var. Inequal.*, vol. 26, no. 2, pp. 110–122, 2023, doi: 10.52783/anvi.v26.i2.287.
- Susanti, E. T., & Sari, H. L. (2020). Pendidikan Kesehatan Tentang Jenis-Jenis Alat Kontrasepsi Terhadap Pemilihan Alat Kontrasepsi. *Jurnal Kesehatan*, 9(1), 53-57.
- Rambe, I. R. (2023). Implementasi Algoritma Apriori Menggunakan Tanagra Pada Coffe Ayos Untuk Mengetahui Pola Penjualan (Doctoral dissertation, Universitas Labuhanbatu).
- Sari, R. M. (2025). PEMANFAATAN KECERDASAN BUATAN (AI) UNTUK Mendukung Program Keluarga Berencana di Indonesia. *Integrative Perspectives of Social and Science Journal*, 2(04 September), 7095-7101.
- NST, M. A. P. (2025). METODE KKN DAN ID3 UNTUK KLASIFIKASI DATA.

Bangun, B., & Karim, A. K. (2024). Pengembalian Data Yang Hilang Pada Dataset Dengan Menggunakan Algoritma K-Nearest Neighbor Imputation Data Mining. *Jurnal Media Informatika Budidarma*, 8(3), 1706.

NURHIKMAH WULANDARI, N. P. M. (2025). KOMPARASI ALGORITMA NAÏVE BAYES DAN KNN DALAM PEMBAGIAN BANTUAN DESA.

PROFIL PENULIS



Euis Sartika merupakan mahasiswa Universitas Labuhanbatu, Fakultas Sains dan Teknologi, Program Studi Sistem Informasi, angkatan 2022, yang lahir di Babussalam pada tanggal 26 April 2004. Penulis menempuh pendidikan pada jenjang sarjana dan memiliki ketertarikan dalam bidang teknologi informasi serta pengembangan sistem berbasis informasi. Penyusunan buku tugas akhir ini merupakan bentuk penerapan ilmu yang diperoleh selama masa perkuliahan sekaligus sebagai syarat penyelesaian studi, dengan harapan karya ini dapat memberikan manfaat dan menjadi referensi bagi mahasiswa maupun pihak lain yang membutuhkan.



Ibnu Rasyid Munthe, S.T., M.Kom. lahir di Kota Rantau Prapat, Kabupaten Labuhanbatu, pada tahun 1987. Beliau memperoleh gelar Sarjana Teknik (S.T.) dari Universitas Nurtanio Bandung dan melanjutkan pendidikan Magister di Universitas Putra Indonesia (YPTK) Padang dengan gelar Magister Komputer (M.Kom.). Beliau aktif dalam menulis artikel ilmiah serta buku akademik. Saat ini, beliau berkiprah sebagai dosen di Universitas Labuhanbatu. Selain itu, beliau juga mengelola kanal YouTube @manjaddawajada2022 sebagai media pembelajaran dan sarana berbagi ilmu pengetahuan.



Marnis Nasution, S.Kom., M.Kom Lahir di Bengkulu 30 maret 1990. Selama sekolah dasar sampai menengah ditempuh di kota Bengkulu. Melanjutkan Pendidikan tinggi strata-1 dan strata-2 di Universitas Putra Indonesia “YPTK” Padang dari tahun 2008 sampai 2024 dengan jurusan Sistem Informasi. Saat ini aktif menjadi Dosen Yayasan di Universitas Labuhanbatu, Sumatera Utara dan menulis beberapa karya Ilmiah dan buku.



Budianto Bangun, S.Sos.,M.Kom. Lahir Medan 04 April 1970 Pendidikan S1 Dharmawangsa Medan dengan gelar (S.Sos), S2 Universitas Putra Indonesia (YPTK) Padang dengan gelar Magister Komputer (M.Kom). Saat ini sebagai tenaga pengajar Universitas Labuhanbatu dan aktif di kegiatan sosial lainnya.

Klasifikasi Pemilihan

Alat Kontrasepsi pada Program Keluarga Berencana

Menggunakan Algoritma
K-Nearest Neighbor



Program Keluarga Berencana (KB) memiliki peran penting dalam meningkatkan kualitas kesehatan keluarga dan pengendalian pertumbuhan penduduk. Namun, pemilihan alat kontrasepsi yang tepat seringkali menjadi tantangan karena dipengaruhi oleh berbagai faktor, seperti usia, jumlah anak, kondisi kesehatan, riwayat penggunaan kontrasepsi, tingkat pendidikan, serta preferensi pasangan. Ketidaktepatan dalam memilih metode kontrasepsi dapat berdampak pada rendahnya efektivitas penggunaan dan meningkatnya risiko efek samping.

Ini bertujuan untuk mengembangkan sistem klasifikasi pemilihan alat kontrasepsi menggunakan algoritma K-Nearest Neighbor (KNN) sebagai metode dalam teknik data mining. Algoritma KNN dipilih karena kemampuannya dalam melakukan klasifikasi berdasarkan kedekatan karakteristik data baru dengan data historis yang telah diketahui kelasnya. Dengan memanfaatkan data peserta KB sebelumnya, sistem akan mengelompokkan dan merekomendasikan jenis kontrasepsi yang paling sesuai, seperti pil, suntik, IUD, implan, atau metode lainnya. Proses penelitian meliputi tahap pengumpulan data, preprocessing, normalisasi, penentuan nilai parameter K, hingga evaluasi akurasi model menggunakan teknik validasi. Hasil yang diperoleh diharapkan mampu memberikan tingkat akurasi yang baik dalam mengklasifikasikan pilihan alat kontrasepsi sesuai dengan karakteristik calon akseptor.

