Journal of Data Science and Scientific Modeling
Vol. 1, No. 1, June 2026, pp. 21-30
eISSN 3163-6829 | https://journal.gemacendekia.org/index.php/jdssm

Pemodelan Epidemiologi Berbasis Data untuk Memprediksi Penyebaran Penyakit Menular di Wilayah Perkotaan

Rizal Mahendra [a,1,*], Putri Ayu Laksmi [a,2]

[a] Program Studi Kesehatan Masyarakat, Fakultas Ilmu Kesehatan, Universitas Kesehatan Digital Indonesia, Indonesia
[1] rizalmahendra@gmail.com*; [2] putriayu@gmail.com
* Corresponding Author

ABSTRACT

Infectious diseases remain a major challenge in public health, particularly in urban areas characterized by high population density, dynamic mobility, and intensive social interactions. These conditions increase the risk of rapid and widespread disease transmission, requiring effective prevention and control strategies. This study aims to develop a data-driven epidemiological model to predict the spread patterns of infectious diseases in urban settings. The approach integrates data analysis techniques with the Susceptible-Infected-Recovered (SIR) model to represent disease transmission dynamics based on population characteristics and environmental factors. The dataset includes daily case counts, population density, mobility levels, vaccination rates, and demographic information obtained from health institutions and public data sources. The research process involves data collection, data cleaning, statistical exploration, model development, and validation using Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). Simulation of various public health intervention scenarios is also conducted to assess policy impacts on transmission rates. The results show that the model provides accurate short- and medium-term predictions. Population density and mobility significantly influence case increases, while vaccination and mobility restrictions effectively reduce active cases. This model supports evidence-based decision-making and early warning systems in urban public health management.

Copyright © 2026, The Author(s)
This is an open-access article under the CC–BY-SA license

Article History
Received 2026-03-22
Revised 2026-04-24
Accepted 2026-04-30
Published 2026-06-23

Keywords
Epidemiologi;
Data Science;
Penyakit Menular;
Model SIR;
Prediksi Penyebaran Penyakit

PENDAHULUAN

Penyakit menular masih menjadi salah satu permasalahan kesehatan masyarakat yang mendapatkan perhatian serius di berbagai negara, termasuk Indonesia. Perkembangan urbanisasi yang pesat telah menyebabkan peningkatan jumlah penduduk di wilayah perkotaan, sehingga menciptakan lingkungan yang sangat rentan terhadap penyebaran berbagai penyakit menular. Tingginya kepadatan penduduk, mobilitas yang intensif, serta interaksi sosial yang tinggi menjadi faktor yang mempercepat proses transmisi penyakit dalam suatu komunitas. Berbagai kejadian wabah yang terjadi dalam beberapa dekade terakhir menunjukkan bahwa wilayah perkotaan sering kali menjadi pusat penyebaran penyakit karena karakteristik sosial dan demografis yang dimilikinya (Kementerian Kesehatan Republik Indonesia, 2023).

Perkembangan penyakit menular tidak hanya berdampak pada aspek kesehatan, tetapi juga memberikan konsekuensi yang luas terhadap sektor ekonomi, pendidikan, dan sosial masyarakat. Ketika suatu wabah terjadi, aktivitas ekonomi dapat mengalami perlambatan akibat pembatasan mobilitas masyarakat, sementara sektor pendidikan menghadapi gangguan proses pembelajaran. Kondisi tersebut menunjukkan bahwa pengendalian penyakit menular memerlukan pendekatan yang komprehensif dan berbasis bukti agar dampak yang ditimbulkan dapat diminimalkan secara efektif (World Health Organization, 2023).

How to cite: Mahendra, R., & Laksmi, P. A. (2026). Pemodelan Epidemiologi Berbasis Data untuk Memprediksi Penyebaran Penyakit Menular di Wilayah Perkotaan. *Journal of Data Science and Scientific Modeling, 1*(1), 21-30. https://doi.org/10.71094/jdssm.v1i1.79

Dalam konteks kesehatan masyarakat modern, pemanfaatan teknologi informasi dan analisis data telah menjadi bagian penting dalam mendukung pengambilan keputusan. Data yang dihasilkan dari sistem surveilans kesehatan, rumah sakit, laboratorium, maupun sumber data digital lainnya dapat dimanfaatkan untuk memahami pola penyebaran penyakit secara lebih mendalam. Melalui pendekatan data science, berbagai informasi yang sebelumnya tersembunyi dalam kumpulan data besar dapat diidentifikasi dan dianalisis untuk menghasilkan prediksi yang akurat mengenai perkembangan suatu penyakit (Han et al., 2022).

Data science merupakan disiplin ilmu yang menggabungkan statistik, komputasi, kecerdasan buatan, dan analisis data untuk menghasilkan informasi yang berguna dalam mendukung proses pengambilan keputusan. Dalam bidang kesehatan, data science telah digunakan untuk memprediksi risiko penyakit, mengidentifikasi faktor-faktor penyebab penyebaran wabah, serta membantu pemerintah dalam merumuskan kebijakan kesehatan yang lebih efektif. Kemampuan data science dalam mengolah data dalam jumlah besar menjadikannya sebagai salah satu pendekatan yang relevan untuk menghadapi tantangan epidemiologi di era digital (Provost & Fawcett, 2013).

Selain data science, pemodelan epidemiologi juga memiliki peran penting dalam memahami dinamika penyebaran penyakit menular. Pemodelan epidemiologi memungkinkan peneliti untuk menggambarkan interaksi antara individu yang rentan, individu yang terinfeksi, dan individu yang telah pulih dari suatu penyakit. Salah satu model yang paling banyak digunakan adalah model Susceptible-Infected-Recovered (SIR), yang mampu menjelaskan proses transmisi penyakit melalui pendekatan matematis yang relatif sederhana namun efektif (Kermack & McKendrick, 1927; Keeling & Rohani, 2011).

Model SIR telah banyak diterapkan dalam berbagai penelitian epidemiologi karena kemampuannya dalam memprediksi jumlah kasus pada periode tertentu. Model ini membagi populasi ke dalam tiga kelompok utama, yaitu individu rentan (susceptible), individu terinfeksi (infected), dan individu sembuh (recovered). Dengan memasukkan parameter tingkat penularan dan tingkat pemulihan, model SIR dapat memberikan gambaran mengenai perkembangan suatu wabah dari waktu ke waktu. Berbagai penelitian menunjukkan bahwa model ini masih relevan digunakan sebagai dasar pengembangan sistem prediksi penyakit menular modern (Brauer et al., 2019; Wearing et al., 2005; Wibowo et al., 2021).

Perkembangan teknologi komputasi telah memungkinkan integrasi antara model epidemiologi dengan teknik analisis data yang lebih canggih. Integrasi tersebut menghasilkan pendekatan pemodelan berbasis data yang tidak hanya mengandalkan asumsi matematis, tetapi juga memanfaatkan data aktual yang diperoleh dari lapangan. Dengan demikian, model yang dihasilkan dapat memberikan hasil prediksi yang lebih akurat dibandingkan dengan pendekatan konvensional. Penggunaan data real time juga memungkinkan sistem untuk beradaptasi terhadap perubahan kondisi epidemiologis yang terjadi secara dinamis (Shmueli & Koppius, 2011; Vespignani, 2012).

Di Indonesia, kebutuhan akan sistem prediksi penyakit menular semakin meningkat seiring dengan munculnya berbagai ancaman penyakit infeksi baru maupun penyakit yang kembali muncul. Tingginya mobilitas penduduk antarwilayah serta pertumbuhan kawasan perkotaan menjadi faktor yang memperbesar risiko penyebaran penyakit secara cepat. Oleh karena itu, pemerintah dan institusi kesehatan memerlukan instrumen yang mampu memberikan peringatan dini sehingga langkah-langkah mitigasi dapat dilakukan sebelum terjadi peningkatan kasus yang signifikan (Badan Pusat Statistik, 2023).

Berbagai penelitian sebelumnya telah menunjukkan bahwa faktor kepadatan penduduk memiliki hubungan yang kuat dengan tingkat penyebaran penyakit menular. Wilayah dengan

kepadatan tinggi cenderung memiliki tingkat kontak antarindividu yang lebih besar sehingga peluang terjadinya transmisi penyakit juga meningkat. Selain itu, mobilitas masyarakat yang tinggi melalui transportasi umum maupun aktivitas ekonomi dapat mempercepat penyebaran penyakit dari satu wilayah ke wilayah lainnya. Oleh karena itu, kedua faktor tersebut perlu diperhatikan dalam pembangunan model prediksi epidemiologi (Rocklöv & Sjödin, 2020).

Selain kepadatan dan mobilitas, tingkat vaksinasi juga menjadi salah satu faktor penting dalam pengendalian penyakit menular. Program vaksinasi yang efektif dapat mengurangi jumlah individu rentan dalam suatu populasi sehingga memperlambat laju penyebaran penyakit. Berbagai studi menunjukkan bahwa peningkatan cakupan vaksinasi berkontribusi terhadap penurunan angka kesakitan dan kematian akibat penyakit menular. Oleh karena itu, variabel vaksinasi perlu diintegrasikan ke dalam model epidemiologi berbasis data untuk meningkatkan akurasi prediksi yang dihasilkan (Anderson & May, 1991).

Meskipun berbagai penelitian mengenai epidemiologi telah dilakukan, masih terdapat keterbatasan dalam pemanfaatan data yang tersedia secara real time untuk mendukung sistem prediksi yang adaptif. Sebagian besar penelitian hanya berfokus pada analisis historis tanpa mengintegrasikan berbagai variabel dinamis yang dapat memengaruhi penyebaran penyakit. Kondisi ini menunjukkan adanya kebutuhan untuk mengembangkan model yang mampu mengombinasikan pendekatan epidemiologi klasik dengan metode data science modern guna menghasilkan prediksi yang lebih komprehensif dan responsif terhadap perubahan situasi lapangan (Lazer et al., 2014; Prasetyo et al., 2021).

Pengembangan model epidemiologi berbasis data menjadi salah satu solusi yang potensial untuk menjawab tantangan tersebut. Dengan memanfaatkan data kasus harian, kepadatan penduduk, mobilitas masyarakat, dan tingkat vaksinasi, model dapat digunakan untuk memprediksi perkembangan penyakit menular secara lebih akurat. Informasi yang dihasilkan dari model tersebut diharapkan dapat membantu pemerintah daerah, tenaga kesehatan, dan pembuat kebijakan dalam menyusun strategi pencegahan dan pengendalian penyakit yang lebih efektif dan tepat sasaran (Nsoesie et al., 2014).

Berdasarkan uraian tersebut, penelitian ini bertujuan untuk mengembangkan model epidemiologi berbasis data yang mampu memprediksi penyebaran penyakit menular di wilayah perkotaan. Penelitian ini mengintegrasikan pendekatan data science dengan model SIR untuk menganalisis hubungan antara faktor-faktor yang memengaruhi penyebaran penyakit dan menghasilkan prediksi yang dapat digunakan sebagai dasar pengambilan keputusan kesehatan masyarakat. Diharapkan hasil penelitian ini dapat memberikan kontribusi terhadap pengembangan sistem surveilans epidemiologi yang lebih modern, adaptif, dan berbasis bukti ilmiah.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan metode pemodelan epidemiologi berbasis data (data-driven epidemiological modeling) untuk memprediksi penyebaran penyakit menular di wilayah perkotaan. Pendekatan kuantitatif dipilih karena memungkinkan pengukuran hubungan antarvariabel secara objektif dan sistematis melalui analisis statistik serta pemodelan matematis yang didukung oleh data empiris. Metode ini telah banyak digunakan dalam penelitian kesehatan masyarakat untuk memahami pola penyebaran penyakit dan mengevaluasi efektivitas berbagai strategi intervensi kesehatan (Ariawan et al., 2021; Kurniawan & Wibowo, 2022).

Penelitian dilaksanakan pada wilayah perkotaan yang memiliki karakteristik kepadatan penduduk tinggi, mobilitas masyarakat yang intensif, serta riwayat kasus penyakit menular yang relatif tinggi dalam beberapa tahun terakhir. Pemilihan lokasi penelitian dilakukan secara purposive berdasarkan ketersediaan data epidemiologi dan data pendukung lainnya yang relevan untuk kebutuhan pemodelan. Pendekatan ini sesuai dengan penelitian epidemiologi modern yang menekankan pentingnya ketersediaan data berkualitas dalam menghasilkan model prediksi yang akurat (Sari et al., 2022; Handayani & Nugroho, 2021).

Data yang digunakan dalam penelitian terdiri atas data sekunder yang diperoleh dari dinas kesehatan daerah, badan pusat statistik, laporan kesehatan nasional, serta sumber data terbuka yang dapat diakses publik. Data tersebut meliputi jumlah kasus harian penyakit menular, jumlah penduduk, kepadatan penduduk, tingkat mobilitas masyarakat, cakupan vaksinasi, tingkat kesembuhan, serta jumlah kematian akibat penyakit yang diteliti. Penggunaan data sekunder dalam penelitian kesehatan masyarakat dinilai efektif untuk menghasilkan analisis yang komprehensif dengan cakupan populasi yang luas (Prasetyo et al., 2021; Hidayat et al., 2023).

Tahapan awal penelitian dilakukan melalui proses pengumpulan dan integrasi data dari berbagai sumber. Data yang diperoleh kemudian diseleksi berdasarkan kelengkapan, konsistensi, dan relevansinya terhadap tujuan penelitian. Selanjutnya dilakukan proses data cleaning untuk menghilangkan data ganda, memperbaiki inkonsistensi format, serta menangani missing value menggunakan teknik imputasi statistik. Tahapan ini sangat penting karena kualitas data akan memengaruhi performa model yang dihasilkan (Widodo & Lestari, 2020; Yuliana et al., 2022).

Setelah proses pembersihan data selesai, dilakukan analisis deskriptif untuk menggambarkan karakteristik data penelitian. Analisis ini mencakup perhitungan nilai rata-rata, median, standar deviasi, nilai minimum, dan maksimum dari setiap variabel penelitian. Selain itu, visualisasi data dilakukan menggunakan grafik tren harian, histogram, dan diagram sebar untuk mengidentifikasi pola distribusi data serta hubungan awal antarvariabel. Analisis deskriptif berfungsi sebagai dasar dalam memahami karakteristik epidemiologi penyakit yang diteliti (Putra & Siregar, 2021).

Tahap berikutnya adalah pembangunan model epidemiologi menggunakan pendekatan Susceptible-Infected-Recovered (SIR). Model SIR dipilih karena mampu menggambarkan dinamika penyebaran penyakit menular melalui tiga kelompok populasi utama, yaitu individu rentan (susceptible), individu terinfeksi (infected), dan individu yang telah sembuh atau memiliki kekebalan (recovered). Model ini telah menjadi salah satu model epidemiologi yang paling banyak digunakan dalam penelitian prediksi penyakit menular karena kesederhanaan dan kemampuannya dalam merepresentasikan proses transmisi penyakit (Rahman et al., 2020; Fauzi et al., 2022).

Secara matematis, model SIR digunakan untuk mengestimasi laju penularan (β) dan laju pemulihan (γ) berdasarkan data aktual yang tersedia. Nilai parameter tersebut diperoleh melalui proses kalibrasi model menggunakan metode optimasi numerik sehingga mampu menghasilkan tingkat kesesuaian yang tinggi terhadap data kasus aktual. Proses estimasi parameter dilakukan dengan memanfaatkan perangkat lunak analisis data yang mendukung komputasi statistik dan pemodelan matematis (Sutrisno & Hakim, 2021).

Selain model epidemiologi konvensional, penelitian ini juga mengintegrasikan pendekatan data science melalui analisis prediktif berbasis machine learning. Metode Random Forest digunakan untuk mengidentifikasi faktor-faktor yang paling berpengaruh terhadap peningkatan jumlah kasus penyakit menular. Variabel independen yang digunakan meliputi kepadatan penduduk, mobilitas masyarakat, cakupan vaksinasi, dan faktor demografis lainnya. Integrasi antara model epidemiologi dan machine learning diharapkan mampu meningkatkan akurasi

prediksi dibandingkan penggunaan satu metode secara terpisah (Mulyani et al., 2023; Han et al., 2022).

Selanjutnya dilakukan pembagian data menjadi data pelatihan (training data) sebesar 80% dan data pengujian (testing data) sebesar 20%. Pembagian ini bertujuan untuk mengukur kemampuan model dalam melakukan prediksi terhadap data yang belum pernah digunakan selama proses pelatihan. Teknik validasi silang (cross-validation) juga diterapkan untuk mengurangi risiko overfitting dan memastikan model memiliki kemampuan generalisasi yang baik pada berbagai kondisi data (Purnama & Dewi, 2022).

Evaluasi performa model dilakukan menggunakan beberapa indikator statistik, yaitu Mean Absolute Error (MAE), Root Mean Square Error (RMSE), dan Mean Absolute Percentage Error (MAPE). Ketiga indikator tersebut dipilih karena mampu memberikan gambaran yang komprehensif mengenai tingkat kesalahan prediksi model. Semakin kecil nilai MAE, RMSE, dan MAPE yang diperoleh, maka semakin baik kemampuan model dalam memprediksi penyebaran penyakit menular (Nugraha et al., 2021).

Untuk mendukung proses pengambilan keputusan, penelitian ini juga melakukan simulasi berbagai skenario intervensi kesehatan masyarakat. Skenario yang diuji meliputi peningkatan cakupan vaksinasi, pengurangan mobilitas masyarakat, peningkatan kepatuhan terhadap protokol kesehatan, serta kombinasi dari beberapa kebijakan tersebut. Simulasi dilakukan dengan mengubah nilai parameter dalam model SIR sehingga dapat diketahui dampak masing-masing intervensi terhadap jumlah kasus aktif, laju penularan, dan durasi epidemi. Pendekatan simulasi semacam ini banyak digunakan dalam penelitian epidemiologi sebagai dasar penyusunan kebijakan kesehatan berbasis bukti (Wahyuni et al., 2022; Permatasari & Yusuf, 2021).

Seluruh proses pengolahan dan analisis data dilakukan menggunakan perangkat lunak Python dengan dukungan pustaka Pandas, NumPy, Scikit-Learn, Matplotlib, dan SciPy. Pemanfaatan perangkat lunak open-source memberikan fleksibilitas dalam pengolahan data berskala besar dan mendukung reproduksibilitas hasil penelitian. Selain itu, dokumentasi kode dan proses analisis dilakukan secara sistematis untuk memastikan transparansi penelitian dan memudahkan proses verifikasi oleh peneliti lain (Utami et al., 2023).

Pada tahap akhir, hasil prediksi dan simulasi yang diperoleh dianalisis secara komprehensif untuk mengidentifikasi pola penyebaran penyakit serta faktor-faktor yang memengaruhinya. Temuan penelitian kemudian dibandingkan dengan kondisi epidemiologi aktual serta hasil penelitian terdahulu guna menilai validitas dan relevansi model yang dikembangkan. Dengan pendekatan ini, penelitian diharapkan mampu menghasilkan model prediksi yang akurat dan dapat digunakan sebagai dasar penyusunan kebijakan kesehatan masyarakat yang lebih efektif dalam mengendalikan penyebaran penyakit menular di wilayah perkotaan (Ariawan et al., 2021; Mulyani et al., 2023).

HASIL DAN PEMBAHASAN

Penelitian ini bertujuan untuk mengembangkan model epidemiologi berbasis data dalam memprediksi penyebaran penyakit menular di wilayah perkotaan dengan mengintegrasikan pendekatan Data Science dan model Susceptible-Infected-Recovered (SIR). Analisis dilakukan menggunakan data kasus harian, kepadatan penduduk, tingkat mobilitas masyarakat, cakupan vaksinasi, dan karakteristik demografis. Hasil pengolahan data menunjukkan bahwa tren penyebaran penyakit memiliki pola fluktuatif yang dipengaruhi oleh tingkat interaksi sosial dan mobilitas penduduk. Wilayah dengan kepadatan penduduk yang tinggi cenderung mengalami

peningkatan jumlah kasus yang lebih cepat dibandingkan wilayah dengan kepadatan rendah. Temuan ini sejalan dengan penelitian yang menyatakan bahwa kepadatan populasi merupakan faktor penting dalam meningkatkan peluang terjadinya kontak antarindividu yang berpotensi mempercepat transmisi penyakit menular (Kurniawan & Hidayat, 2022).

Pada tahap eksplorasi data, ditemukan adanya korelasi positif antara tingkat mobilitas masyarakat dan jumlah kasus harian. Ketika aktivitas masyarakat meningkat, terutama pada kawasan pusat ekonomi dan transportasi, jumlah kasus baru juga menunjukkan kecenderungan meningkat. Analisis korelasi menghasilkan nilai koefisien yang menunjukkan hubungan sedang hingga kuat antara kedua variabel tersebut. Hasil ini mengindikasikan bahwa mobilitas menjadi salah satu parameter penting yang perlu diperhitungkan dalam pengembangan model prediksi epidemiologi. Temuan ini mendukung penelitian sebelumnya yang menunjukkan bahwa pergerakan penduduk berkontribusi signifikan terhadap penyebaran penyakit menular di kawasan perkotaan (Prasetyo et al., 2021).

Model SIR yang dikembangkan dalam penelitian ini berhasil menggambarkan dinamika penyebaran penyakit secara cukup akurat. Nilai parameter transmisi (β) dan tingkat pemulihan (γ) diperoleh melalui proses estimasi menggunakan data historis kasus. Berdasarkan hasil simulasi, fase peningkatan jumlah kasus terjadi pada periode awal penyebaran penyakit, diikuti oleh fase puncak epidemi dan fase penurunan kasus setelah berbagai intervensi kesehatan diterapkan. Pola tersebut menunjukkan kesesuaian dengan teori epidemiologi yang menjelaskan bahwa jumlah individu rentan akan berkurang seiring meningkatnya jumlah individu yang terinfeksi dan sembuh.

Evaluasi performa model dilakukan menggunakan indikator Mean Absolute Error (MAE) dan Root Mean Square Error (RMSE). Hasil pengujian menunjukkan bahwa model memiliki tingkat kesalahan prediksi yang relatif rendah. Nilai MAE dan RMSE yang diperoleh menunjukkan bahwa model mampu mengikuti tren aktual penyebaran penyakit dengan tingkat akurasi yang baik. Keakuratan model ini membuktikan bahwa kombinasi pendekatan statistik dan machine learning sederhana dengan model epidemiologi mampu menghasilkan prediksi yang dapat digunakan dalam proses perencanaan kebijakan kesehatan masyarakat.

Selain mengukur akurasi prediksi, penelitian ini juga melakukan simulasi skenario intervensi kesehatan. Pada skenario pertama, peningkatan cakupan vaksinasi sebesar 20% menunjukkan penurunan jumlah kasus aktif secara signifikan dibandingkan kondisi dasar. Simulasi memperlihatkan bahwa semakin tinggi tingkat vaksinasi, semakin kecil jumlah individu rentan yang berpotensi terinfeksi. Hasil ini memperkuat temuan berbagai penelitian yang menyatakan bahwa vaksinasi merupakan strategi paling efektif dalam menekan laju penyebaran penyakit menular di tingkat populasi (Rahman & Sari, 2023).

Skenario kedua menguji dampak pembatasan mobilitas masyarakat terhadap laju penyebaran penyakit. Hasil simulasi menunjukkan bahwa pengurangan mobilitas sebesar 30% mampu memperlambat peningkatan jumlah kasus harian dan menunda terjadinya puncak epidemi. Penundaan tersebut memberikan kesempatan bagi fasilitas kesehatan untuk mempersiapkan sumber daya yang diperlukan dalam menghadapi lonjakan pasien. Dengan demikian, kebijakan pengendalian mobilitas dapat menjadi instrumen yang efektif dalam mengurangi tekanan terhadap sistem pelayanan kesehatan.

Hasil penelitian juga menunjukkan bahwa faktor demografis memiliki pengaruh terhadap variasi penyebaran penyakit. Kelompok usia produktif menunjukkan tingkat risiko penularan yang lebih tinggi dibandingkan kelompok usia lainnya karena memiliki frekuensi aktivitas dan interaksi sosial yang lebih besar. Selain itu, wilayah dengan tingkat kepadatan hunian yang tinggi cenderung mengalami penyebaran penyakit yang lebih cepat. Temuan ini menunjukkan bahwa

strategi mitigasi perlu disesuaikan dengan karakteristik populasi agar lebih efektif dalam mengendalikan penyebaran penyakit.

Dari perspektif data science, penggunaan data historis dan data real-time memberikan keuntungan dalam meningkatkan kemampuan model untuk melakukan prediksi secara dinamis. Model dapat diperbarui secara berkala sesuai dengan perubahan kondisi lapangan sehingga menghasilkan informasi yang lebih relevan bagi pengambil kebijakan. Pendekatan ini sejalan dengan perkembangan sistem kesehatan digital yang menekankan pemanfaatan big data dan analitik prediktif untuk mendukung pengambilan keputusan berbasis bukti (Utami et al., 2022).

Meskipun model yang dikembangkan menunjukkan performa yang baik, terdapat beberapa keterbatasan yang perlu diperhatikan. Kualitas data menjadi faktor utama yang memengaruhi hasil prediksi. Ketidaklengkapan data pelaporan kasus, keterlambatan pencatatan, serta variasi metode pengumpulan data dapat menyebabkan bias dalam proses pemodelan. Selain itu, model SIR memiliki asumsi bahwa populasi bersifat homogen, padahal dalam kenyataannya terdapat perbedaan karakteristik sosial, ekonomi, dan geografis yang dapat memengaruhi pola penyebaran penyakit.

Secara keseluruhan, hasil penelitian menunjukkan bahwa integrasi antara model epidemiologi dan teknik data science mampu menghasilkan sistem prediksi yang efektif untuk memantau dan memperkirakan penyebaran penyakit menular di wilayah perkotaan. Kemampuan model dalam mensimulasikan berbagai skenario kebijakan memberikan nilai tambah bagi pemerintah dan instansi kesehatan dalam menyusun strategi pencegahan yang lebih terukur dan berbasis data. Oleh karena itu, pengembangan model epidemiologi berbasis data dapat menjadi salah satu solusi inovatif dalam meningkatkan kesiapsiagaan menghadapi wabah penyakit di masa mendatang.

Lebih lanjut, implementasi model ini berpotensi mendukung pembangunan sistem peringatan dini (early warning system) yang mampu mendeteksi peningkatan risiko penyebaran penyakit sebelum terjadi lonjakan kasus yang signifikan. Dengan adanya prediksi yang lebih akurat, pemerintah daerah dapat mengalokasikan sumber daya kesehatan secara lebih efisien, termasuk distribusi tenaga medis, obat-obatan, dan fasilitas pelayanan kesehatan. Hal ini menunjukkan bahwa pemanfaatan teknologi analitik dan pemodelan ilmiah tidak hanya berkontribusi pada aspek akademis, tetapi juga memiliki dampak praktis yang besar dalam mendukung ketahanan sistem kesehatan masyarakat di wilayah perkotaan modern.

KESIMPULAN

Penelitian mengenai pemodelan epidemiologi berbasis data untuk memprediksi penyebaran penyakit menular di wilayah perkotaan menunjukkan bahwa integrasi antara metode data science dan model epidemiologi SIR mampu menghasilkan prediksi yang akurat terhadap dinamika penyebaran penyakit. Pemanfaatan data kasus harian, kepadatan penduduk, mobilitas masyarakat, tingkat vaksinasi, dan faktor demografis memberikan gambaran yang lebih komprehensif mengenai pola transmisi penyakit dalam suatu populasi. Hasil penelitian membuktikan bahwa pendekatan berbasis data dapat meningkatkan kemampuan dalam mengidentifikasi tren penyebaran penyakit secara lebih cepat dan tepat.

Temuan penelitian juga menunjukkan bahwa kepadatan penduduk dan mobilitas masyarakat merupakan faktor dominan yang memengaruhi peningkatan jumlah kasus penyakit menular di wilayah perkotaan. Tingginya interaksi sosial dan pergerakan penduduk mempercepat proses transmisi penyakit, sehingga diperlukan strategi pengendalian yang mempertimbangkan kedua aspek tersebut. Selain itu, peningkatan cakupan vaksinasi terbukti

mampu menekan laju penyebaran penyakit dan mengurangi jumlah kasus aktif secara signifikan, terutama ketika diterapkan secara konsisten dan terintegrasi dengan kebijakan kesehatan lainnya.

Berdasarkan hasil simulasi yang dilakukan, model yang dikembangkan dapat digunakan sebagai alat bantu dalam mengevaluasi berbagai skenario intervensi kesehatan masyarakat sebelum kebijakan diterapkan di lapangan. Kemampuan model dalam memberikan estimasi dampak dari pembatasan mobilitas, peningkatan vaksinasi, maupun strategi mitigasi lainnya menjadi nilai tambah yang penting bagi pengambil kebijakan. Dengan demikian, model epidemiologi berbasis data tidak hanya berfungsi sebagai sarana prediksi, tetapi juga sebagai instrumen pendukung perencanaan dan pengambilan keputusan yang berbasis bukti ilmiah.

Secara keseluruhan, penelitian ini menegaskan bahwa pemodelan epidemiologi berbasis data memiliki potensi besar untuk meningkatkan efektivitas sistem peringatan dini dan kesiapsiagaan kesehatan masyarakat di wilayah perkotaan. Pengembangan model yang adaptif terhadap perubahan data secara real time dapat membantu pemerintah dan instansi kesehatan dalam merespons ancaman penyakit menular dengan lebih cepat dan tepat sasaran. Oleh karena itu, penerapan teknologi data science dalam bidang epidemiologi perlu terus dikembangkan guna mendukung terciptanya sistem kesehatan yang lebih tangguh, responsif, dan berkelanjutan.

DAFTAR PUSTAKA

Anderson, R. M., & May, R. M. (1991). *Infectious diseases of humans: Dynamics and control*. Oxford University Press.

Ariawan, I., Putri, D. A., & Santoso, H. (2021). Epidemiological modeling for infectious disease surveillance in urban areas. *Jurnal Kesehatan Masyarakat Indonesia, 16*(2), 112–121.

Badan Pusat Statistik. (2023). *Statistik Indonesia 2023*. Badan Pusat Statistik.

Brauer, F., Castillo-Chavez, C., & Feng, Z. (2019). *Mathematical models in epidemiology*. Springer.

Fauzi, M., Nugroho, A., & Lestari, D. (2022). Application of SIR model in infectious disease prediction. *Jurnal Biostatistika Indonesia, 11*(1), 45–56.

Han, J., Kamber, M., & Pei, J. (2022). *Data mining: Concepts and techniques* (4th ed.). Morgan Kaufmann.

Handayani, S., & Nugroho, R. (2021). Data quality assessment in public health research. *Jurnal Informatika Kesehatan, 9*(3), 88–97.

Hidayat, T., Rahmawati, N., & Kurniawan, A. (2023). Secondary data utilization in epidemiological studies. *Jurnal Statistik Terapan, 14*(1), 23–35.

Keeling, M. J., & Rohani, P. (2011). *Modeling infectious diseases in humans and animals*. Princeton University Press.

Kementerian Kesehatan Republik Indonesia. (2023). *Profil kesehatan Indonesia 2023*. Kementerian Kesehatan RI.

Kermack, W. O., & McKendrick, A. G. (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society A, 115*(772), 700–721. https://doi.org/10.1098/rspa.1927.0118

Kurniawan, A., & Hidayat, R. (2022). Analisis pengaruh kepadatan penduduk terhadap penyebaran penyakit menular di kawasan perkotaan Indonesia. *Jurnal Kesehatan Masyarakat Indonesia, 17*(2), 115–124.

Kurniawan, B., & Wibowo, A. (2022). Quantitative methods for disease outbreak prediction. *Jurnal Ilmu Kesehatan, 18*(2), 75–86.

Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014). The parable of Google Flu: Traps in big data analysis. *Science, 343*(6176), 1203–1205. https://doi.org/10.1126/science.1248506

Mulyani, R., Pratama, D., & Sari, M. (2023). Integration of machine learning and epidemiological models for disease forecasting. *Jurnal Sains Data Indonesia, 5*(1), 34–48.

Nsoesie, E. O., Brownstein, J. S., Ramakrishnan, N., & Marathe, M. V. (2014). A systematic review of studies on forecasting the dynamics of influenza outbreaks. *Influenza and Other Respiratory Viruses, 8*(3), 309–316. https://doi.org/10.1111/irv.12226

Nugraha, P., Setyawan, A., & Yuniarti, S. (2021). Evaluation metrics for predictive health models. *Jurnal Teknologi Informasi dan Kesehatan, 12*(2), 90–102.

Permatasari, D., & Yusuf, M. (2021). Public health intervention simulation using epidemiological approaches. *Jurnal Kebijakan Kesehatan Indonesia, 10*(4), 201–214.

Prasetyo, D., Nugroho, B., & Santoso, A. (2021). Mobilitas penduduk dan dinamika penyebaran penyakit menular berbasis data spasial. *Jurnal Sistem Informasi Kesehatan, 9*(1), 45–57.

Prasetyo, E., Hadi, S., & Lestari, P. (2021). Big data applications in health surveillance systems. *Jurnal Sistem Informasi Kesehatan, 13*(1), 55–67.

Provost, F., & Fawcett, T. (2013). *Data science for business*. O'Reilly Media.

Purnama, R., & Dewi, N. (2022). Cross-validation techniques in predictive analytics. *Jurnal Informatika dan Komputasi, 8*(2), 99–108.

Putra, Y., & Siregar, F. (2021). Descriptive statistical analysis for epidemiological research. *Jurnal Matematika dan Sains, 15*(3), 145–154.

Rahman, F., & Sari, N. P. (2023). Efektivitas program vaksinasi dalam menurunkan angka kejadian penyakit menular di Indonesia. *Jurnal Epidemiologi dan Kesehatan Komunitas, 8*(1), 21–34.

Rahman, M., Saputra, H., & Kholid, A. (2020). Mathematical modeling of infectious diseases using SIR framework. *Jurnal Matematika Terapan, 9*(1), 1–12.

Rocklöv, J., & Sjödin, H. (2020). High population densities catalyse the spread of COVID-19. *Journal of Travel Medicine, 27*(3), taaa038. https://doi.org/10.1093/jtm/taaa038

Sari, N., Wijaya, T., & Hapsari, D. (2022). Urban population density and disease transmission analysis. *Jurnal Kesehatan Lingkungan Indonesia, 21*(1), 41–52.

Shmueli, G., & Koppius, O. R. (2011). Predictive analytics in information systems research. *MIS Quarterly, 35*(3), 553–572. https://doi.org/10.2307/23042796

Sutrisno, A., & Hakim, L. (2021). Parameter estimation methods in epidemiological models. *Jurnal Komputasi Saintifik, 7*(2), 61–73.

Utami, L., Wijaya, M., & Saputra, H. (2022). Pemanfaatan big data analytics untuk mendukung sistem kesehatan digital di Indonesia. *Jurnal Teknologi Informasi dan Kesehatan, 10*(3), 201–214.

Utami, R., Wulandari, S., & Prakoso, A. (2023). Open-source software utilization in scientific data analysis. *Jurnal Teknologi dan Sistem Informasi, 19*(1), 15–27.

Vespignani, A. (2012). Modelling dynamical processes in complex socio-technical systems. *Nature Physics, 8*(1), 32–39. https://doi.org/10.1038/nphys2160

Wahyuni, D., Hasanah, U., & Rizki, M. (2022). Scenario-based modeling for epidemic control strategies. *Jurnal Administrasi Kesehatan Indonesia, 10*(2), 119–132.

Wearing, H. J., Rohani, P., & Keeling, M. J. (2005). Appropriate models for the management of infectious diseases. *PLoS Medicine, 2*(7), e174. https://doi.org/10.1371/journal.pmed.0020174

Wibowo, T., Handayani, S., & Putra, I. (2021). Model epidemiologi SIR sebagai alat prediksi penyebaran penyakit menular pada wilayah perkotaan. *Jurnal Matematika dan Sains Terapan, 15*(4), 289–301.

Widodo, B., & Lestari, E. (2020). Data preprocessing techniques in health analytics. *Jurnal Informatika Medis, 6*(1), 27–39.

World Health Organization. (2023). *Global health observatory: Infectious diseases surveillance report*. World Health Organization.

Yuliana, N., Prasetyo, H., & Andriani, R. (2022). Missing value treatment in public health datasets. *Jurnal Data dan Informasi Kesehatan, 4*(2), 70–81.