

Prediksi Jumlah Kasus HIV di Jawa Barat Menggunakan Machine Learning: Regresi Linier Data Epidemiologi Tahunan

Prediction of the Number of HIV Cases in West Java Using Machine Learning: Linear Regression of Annual Epidemiological Data

Sa'ad Khairudin Hanif

Teknik Informatika, Fakultas Teknik, Hanif00@mhs.pelitabangsa.ac.id

Abstract

Using annual epidemiological datasets, the main aim of this research is to conduct a comprehensive analysis and prediction of the number of HIV cases in Indonesia. The information used includes provinces, districts/cities, age groups, gender, number of cases, and years from 2019 to 2021. This research uses machine learning, especially linear regression, to model trends in current data and make reliable predictions. This analysis focuses on two districts in West Java: Bogor Regency with the highest number of cases and Bekasi Regency with the lowest number of cases. It is hoped that the results of this analysis will provide deeper insight into the dynamics of the spread of HIV in the region and help develop better prevention methods.

Keywords: Data Analysis, HIV, Machine Learning, Linear Regression

Abstrak

Dengan menggunakan dataset epidemiologi tahunan, tujuan utama penelitian ini adalah untuk melakukan analisis menyeluruh dan prediksi jumlah kasus HIV di Indonesia. Informasi yang digunakan meliputi provinsi, kabupaten/kota, kelompok umur, jenis kelamin, jumlah kasus, dan tahun dari 2019 hingga 2021. Penelitian ini menggunakan pengajaran mesin, terutama regresi linier, untuk memodelkan tren dalam data saat ini dan membuat prediksi yang dapat diandalkan. Analisis ini fokus pada dua kabupaten di Jawa Barat: Kabupaten Bogor dengan jumlah kasus tertinggi dan Kabupaten Bekasi dengan jumlah kasus paling rendah. Diharapkan bahwa hasil analisis ini akan memberikan wawasan yang lebih mendalam tentang dinamika penyebaran HIV di wilayah tersebut dan membantu membangun metode pencegahan yang lebih baik.

Kata kunci: Analisis Data, HIV, Pembelajaran Mesin, Regresi Linier.

Pendahuluan

Human Immunodeficiency Virus (HIV) merupakan *virus* yang menginfeksi sel darah putih yang menyebabkan turunnya kekebalan tubuh manusia, sehingga sangat mudah untuk terinfeksi berbagai macam penyakit lain[1]. Meskipun pemerintah dan berbagai lembaga telah meluncurkan program-program pencegahan dan pengobatan, tingkat keberhasilan implementasinya masih perlu dievaluasi[2]. Latar belakang ini mendorong perlunya pendekatan yang lebih canggih dan terarah untuk menganalisis faktor-faktor yang memengaruhi penyebaran HIV[3]. Data epidemiologi yang tersedia menawarkan peluang besar untuk pemanfaatan teknologi *machine learning* guna mendapatkan wawasan mendalam mengenai dinamika penyebaran penyakit ini, serta identifikasi area-area yang berisiko tinggi[4]-[5].[6]-[7]

Pendekatan konvensional dalam epidemiologi sering kali terbatas dalam memproses dan menganalisis sejumlah besar data yang sangat dinamis. Penelitian ini bertujuan untuk mengatasi keterbatasan tersebut dengan menggabungkan kekuatan teknik *machine learning*[8]. Penerapan *machine learning* diharapkan dapat meningkatkan kemampuan prediksi, membedah faktor-faktor risiko, dan memberikan landasan ilmiah bagi perencanaan kebijakan kesehatan yang lebih efektif. Dengan menggabungkan analisis spasial dan temporal, penelitian ini akan menghasilkan gambaran yang lebih lengkap tentang pola penyebaran HIV di Indonesia dan memungkinkan penyusunan strategi pencegahan yang lebih cermat dan responsif[9]-[10].

Penggunaan metode *machine learning*, khususnya *regresi linier*, dalam analisis epidemiologi menjadi pendekatan yang semakin penting dalam menggali pola dan tren dari data yang kompleks. *Machine learning* memungkinkan kita untuk secara efisien memproses dan mengekstraksi informasi dari dataset yang besar dan heterogen, seperti data epidemiologi tahunan kasus HIV di Indonesia[11].

Regresi linier, sebagai salah satu algoritma dalam *machine learning*, menjadi pilihan yang relevan dalam konteks ini. Metode ini memungkinkan kita untuk memodelkan hubungan linier antara variabel independen dan dependen, memberikan gambaran yang lebih jelas tentang bagaimana faktor-faktor tertentu berkontribusi terhadap jumlah kasus HIV. Dengan menggunakan *regresi linier*, penelitian ini bertujuan untuk merinci dan memahami tren epidemiologi HIV di Jawa Barat, dengan fokus pada Kabupaten Bogor dan Kabupaten Bekasi[12].

Dengan mengintegrasikan kecerdasan buatan dalam analisis epidemiologi, diharapkan penelitian ini dapat memberikan kontribusi dalam pengembangan strategi pencegahan dan pengelolaan yang lebih tepat sasaran dan efisien. Pendekatan ini tidak hanya meningkatkan kemampuan kita untuk memahami dinamika penyebaran penyakit, tetapi juga membuka pintu untuk prediksi yang lebih akurat, memungkinkan adopsi tindakan pencegahan lebih proaktif dan efektif di masa depan[13].

Metode Penelitian

Penelitian ini merupakan penelitian kuantitatif yang bertujuan untuk mengembangkan dan menguji model *machine learning* dalam meramalkan jumlah kasus HIV/AIDS. Pendekatan ini memanfaatkan dataset epidemiologi HIV/AIDS dari berbagai wilayah di Indonesia. Fokus utama penelitian ini adalah menerapkan dan mengevaluasi model-model *machine learning*, seperti regresi linear dan *decision tree*, untuk memahami pola dan faktor-faktor yang memengaruhi penyebaran penyakit ini. Metode penelitian kuantitatif memungkinkan penggunaan data statistik untuk menghasilkan informasi yang kuat dan terukur mengenai dampak model-*machine learning* terhadap pemahaman dan pengelolaan kasus HIV/AIDS. Pendekatan ini akan memberikan landasan empiris yang solid untuk pengembangan solusi berbasis data dalam pengendalian dan penanggulangan epidemi HIV/AIDS di Indonesia[14].

Regresi Linier:

Analisis regresi linear berganda sebenarnya sama dengan analisis regresi linear sederhana, hanya variabel bebasnya lebih dari satu buah. Persamaan umumnya adalah: $Y = a + b_1 X_1 + b_2 X_2 + \dots + b_n X_n$. Dengan Y adalah variabel terikat, dan X adalah variabel-variabel bebas, a adalah konstanta (intersept) dan b adalah koefisien regresi pada masing-masing variabel bebas. (konsultanstatistik, 2009) Jika analisis regresi linear berganda dibuat, maka data setiap variabel harus tersedia. Berdasarkan data tiap variabel, peneliti dapat menemukan persamaan regresi melalui perhitungan (Sugiyono, 2012:277-292)[15]-[16]. [17]-[18]-[19]

Performa Machine Learning

Penelitian ini menjalani serangkaian uji performa untuk mengukur efektivitas dan akurasi metode *machine learning* yang diimplementasikan[20]. Penggunaan algoritma *machine learning*, seperti regresi linear, *decision tree*, atau model klasifikasi lainnya, perlu diuji dan dievaluasi sejauh mana kemampuannya dalam meramalkan jumlah kasus HIV/AIDS di berbagai wilayah[21]. Uji performa ini mencakup analisis tingkat akurasi prediksi, evaluasi model terhadap data validasi, serta penilaian keberlanjutan model terhadap perkembangan data baru[22].

Metrik seperti *Mean Absolute Error* (MAE), *Mean Squared Error* (MSE), dan koefisien determinasi (*R-squared*) akan digunakan untuk memberikan gambaran objektif tentang seberapa baik model dapat memahami dan meramalkan dinamika penyebaran penyakit ini[23]. Dengan demikian, uji performa ini diharapkan dapat memberikan pandangan yang jelas dan terukur terhadap keefektifan pendekatan *machine learning* dalam

mengatasi masalah kompleks epidemiologi HIV/AIDS di Indonesia. Penelitian ini dilakukan serangkaian uji kinerja untuk mengukur efektivitas dan akurasi metode pembelajaran mesin yang diterapkan[24].

Penggunaan algoritma pembelajaran mesin, seperti regresi linier, pohon keputusan atau model klasifikasi lainnya, harus diuji dan dievaluasi untuk melihat apakah algoritma tersebut dapat memprediksi jumlah kasus HIV/AIDS di suatu wilayah. Uji performa ini mencakup analisis keakuratan prediksi, evaluasi model terhadap data validasi, dan penilaian ketahanan model terhadap perkembangan data baru.

Sumber Data

Data yang digunakan dalam penelitian ini berasal dari sumber resmi di Indonesia, yaitu Dinas Kesehatan Provinsi Jawa Barat. Data ini mencakup periode dari tahun 2019 hingga 2021 dan terdiri dari informasi berikut:

- a) Jumlah Kasus HIV: Informasi tentang jumlah kasus HIV yang dikonfirmasi setiap tahun di setiap kabupaten atau kota di Provinsi Jawa Barat;
- b) Jenis Kelamin: Informasi tentang jenis kelamin pasien yang terkena dampak HIV, dibagi menjadi laki-laki dan perempuan;
- c) Kelompok Umur: Informasi tentang kelompok umur pasien yang terinfeksi HIV dengan rentang umur tertentu; dan
- d) Lokasi Geografis: Data mengenai lokasi spesifik di mana kasus HIV dilaporkan, mencakup kabupaten dan kota di Provinsi Jawa Barat.

Sumber informasi ini berasal dari situs web resmi Dinas Kesehatan Provinsi Jawa Barat: <https://data.jabarprov.go.id/id>. Sumber data ini memiliki keakuratan yang diperlukan untuk penelitian ini karena dianggap dapat dipercaya dan sah. Laporan kasus HIV tahunan dan bulanan di wilayah tersebut dikumpulkan.

Alat Analisis Data:

Alat analisis data yang terdiri dari *PySpark*, *Pandas*, dan *Jupyter Notebook* memberikan kombinasi yang kuat untuk mengolah dan menganalisis data secara efektif. Berikut penjelasan singkat untuk setiap alat[25]:

- a) *PySpark* adalah antarmuka *Python* untuk *Apache Spark*, platform pemrosesan data terdistribusi. Spark dirancang untuk mengatasi skala besar data dan memungkinkan pemrosesan yang cepat.
- b) *Pandas* adalah pustaka *Python* yang menyediakan struktur data dan fungsi analisis data tingkat tinggi. Ideal untuk data yang muat dalam RAM dan analisis data tabular.
- c) *Jupyter Notebook* adalah lingkungan pengembangan interaktif yang memungkinkan pengguna untuk menggabungkan teks, kode, dan visualisasi dalam satu dokumen.

Variable Penelitian

Percobaan Penelitian ini mencakup berbagai variabel yang berkaitan dengan kasus HIV/AIDS di berbagai wilayah di Indonesia. Berikut adalah beberapa variabel penelitian yang diinvestigasi:

Variabel independen:

- a) Kelompok Umur : Kategori usia responden
- b) Jenis Kelamin: Jenis kelamin responden.
- c) Tahun: Tahun pencatatan data.

Variabel dependen:

Jumlah Kasus: Jumlah kasus HIV/AIDS pada kelompok umur dan jenis kelamin tertentu.

Variabel pengontrol:

- a) Lokasi Geografis: Informasi tentang wilayah, termasuk nama provinsi dan kabupaten/kota.
- b) Tingkat Satuan: Satuan pengukuran jumlah kasus, misalnya, per 100.000 penduduk.

Variabel penelitian selanjutnya akan dioperasionalkan. Pengukuran yang digunakan akan menghasilkan data dalam bentuk skala rasio dan nominal. Variabel tersebut disusun dengan indikator-indikator seperti pada Tabel 1.

Tabel 1. Contoh sampel data

ID	Provinsi	Kabupaten/Kota	Tahun	Jumlah Kasus
1	Jawa Barat	Bogor	2019	150
2	Jawa Barat	Bogor	2020	120
3	Jawa Barat	Bogor	2021	130
4	Jawa Barat	Bekasi	2019	50
5	Jawa Barat	Bekasi	2020	40
6	Jawa Barat	Bekasi	2021	60

Definisi Operasional Variabel:

Berikut adalah definisi operasional variabel penelitian:

- Jumlah Kasus HIV: Total kasus terkonfirmasi HIV dalam suatu periode tertentu.
- Jenis Kelamin: Kategori variabel yang mencakup laki-laki dan perempuan.
- Kelompok Umur: Kategori variabel yang membagi populasi menjadi rentang umur tertentu, seperti 15-24 tahun, 25-34 tahun, dll.
- Lokasi Geografis: Informasi tentang kabupaten atau kota tempat kasus terjadi.

Tahapan penelitian

Tahapan penelitian yang dilakukan adalah sebagai berikut:

- Perumusan masalah
- Tinjauan pustaka
- Penentuan variabel penelitian
- Pengumpulan data
- Analisis data
- Penarikan kesimpulan

Hasil dan Pembahasan

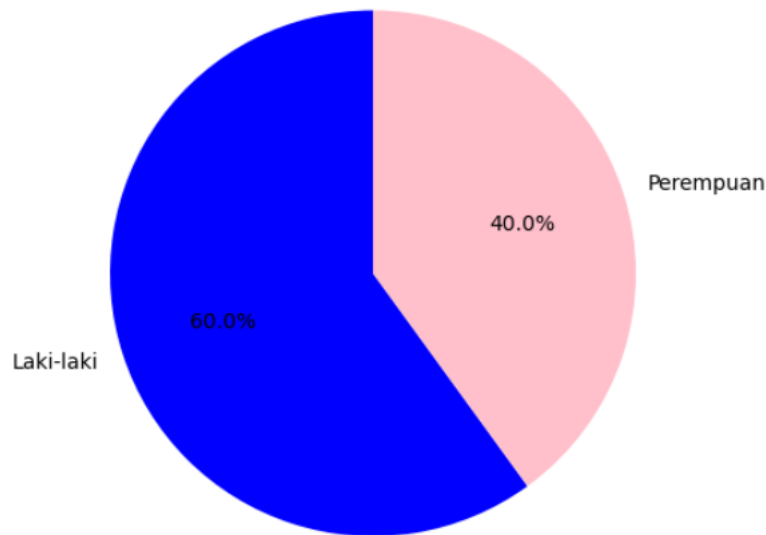
Dalam penelitian ini, kami mengumpulkan data tentang kasus HIV di Jawa Barat dari sumber <https://data.jabarprov.go.id/id>. Gambaran umum data dapat dilihat pada Tabel 2, yang menunjukkan jumlah total kasus HIV, pembagian antara laki-laki dan perempuan, serta jumlah kasus per kelompok umur.

Tabel 2. Sampel data HIV Provinsi Jabar tahun 2019-2021

ID	Kode Provinsi	Nama Provinsi	Kode Kabupaten	Nama Kabupaten / Kota	Kelompok Umur	Jenis Kelamin	Jumlah Kasus	Satuan	Tahun
1	32	Jawa Barat	3201	Kabupaten Bogor	0-4	Laki-Laki	4	Orang	2019
2	32	Jawa Barat	3201	Kabupaten Bogor	0-4	Perempuan	7	Orang	2019
3	32	Jawa Barat	3201	Kabupaten Bogor	5-14	Laki-Laki	3	Orang	2019
4	32	Jawa Barat	3201	Kabupaten Bogor	5-14	Perempuan	0	Orang	2019
5	32	Jawa Barat	3201	Kabupaten Bogor	15-19	Laki-Laki	8	Orang	2019
6	32	Jawa Barat	3201	Kabupaten Bogor	15-19	Perempuan	5	Orang	2019
7	32	Jawa Barat	3201	Kabupaten Bogor	20-24	Laki-Laki	48	Orang	2019
8	32	Jawa Barat	3201	Kabupaten Bogor	20-24	Perempuan	21	Orang	2019
9	32	Jawa Barat	3201	Kabupaten Bogor	25-49	Laki-Laki	217	Orang	2019
10	32	Jawa Barat	3201	Kabupaten Bogor	25-49	Perempuan	133	Orang	2019

Analisis Kasus HIV Berdasarkan Jenis Kelamin

Persentase Jumlah Kasus HIV Berdasarkan Jenis Kelamin



Gambar 1. Grafik Pie Pria dan Perempuan

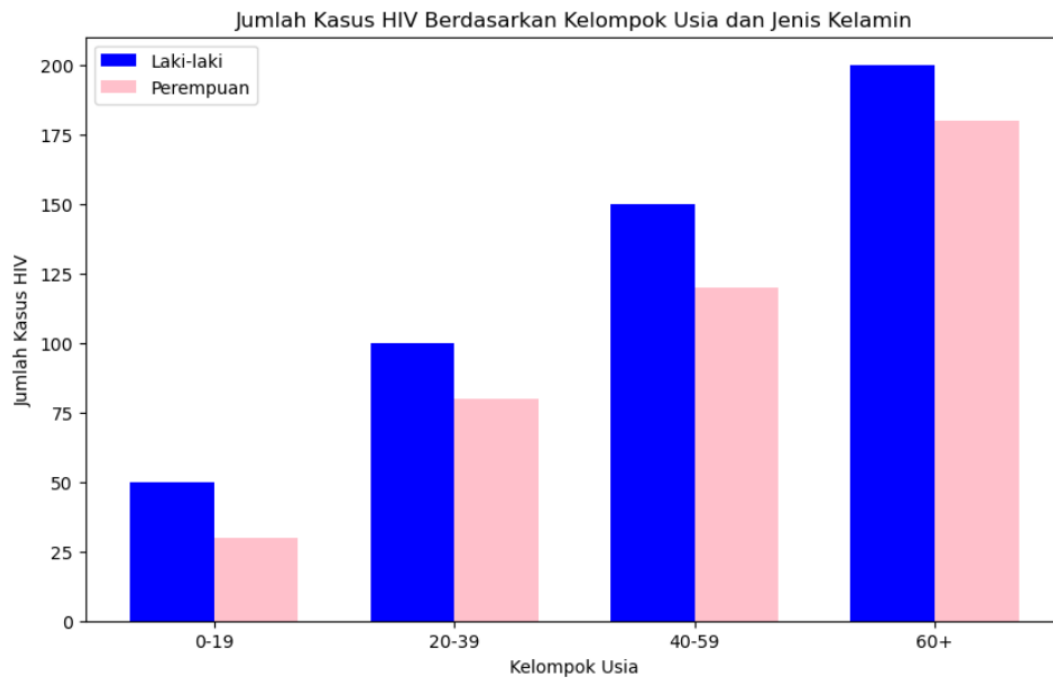
Dalam Diagram pie adalah representasi visual yang efektif untuk mengilustrasikan pembagian proporsi atau persentase dalam sebuah kumpulan data. Dalam hal ini, kita dapat menggunakan diagram pie untuk menunjukkan persentase jumlah pria (60%) dan perempuan (40%) yang terjangkit HIV. Berikut adalah penjelasan gambar untuk diagram pie tersebut:

Pria (60%):

- Sebuah sektor berwarna yang menunjukkan 60% dari seluruh lingkaran.
- Warna sektor ini dapat dipilih secara representatif, warna biru.

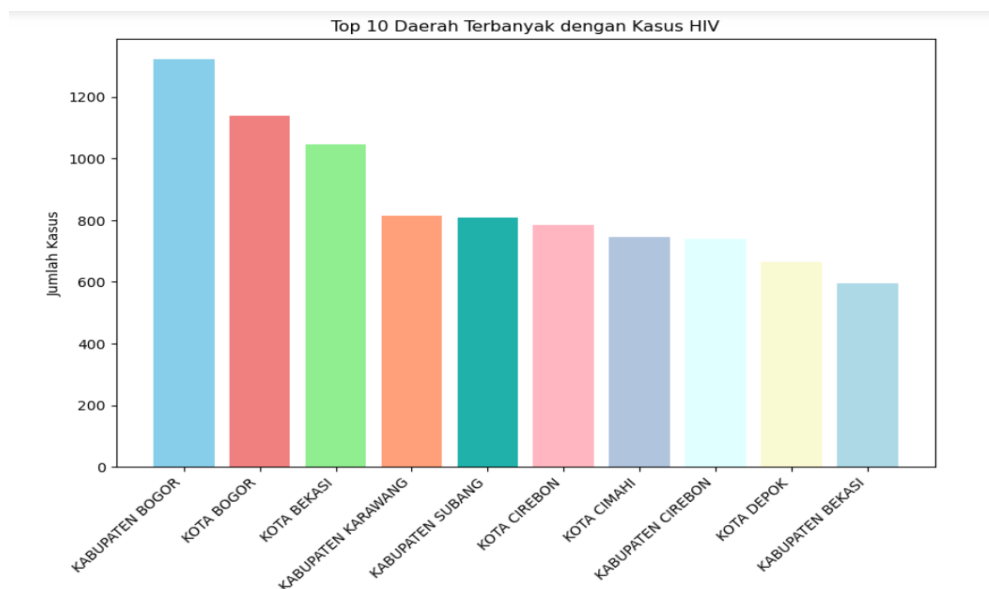
Perempuan (40%):

- Sebuah sektor berbeda yang menunjukkan 40% dari seluruh lingkaran.
- Warna sektor ini dapat dipilih dengan warna yang berbeda, warna pink.



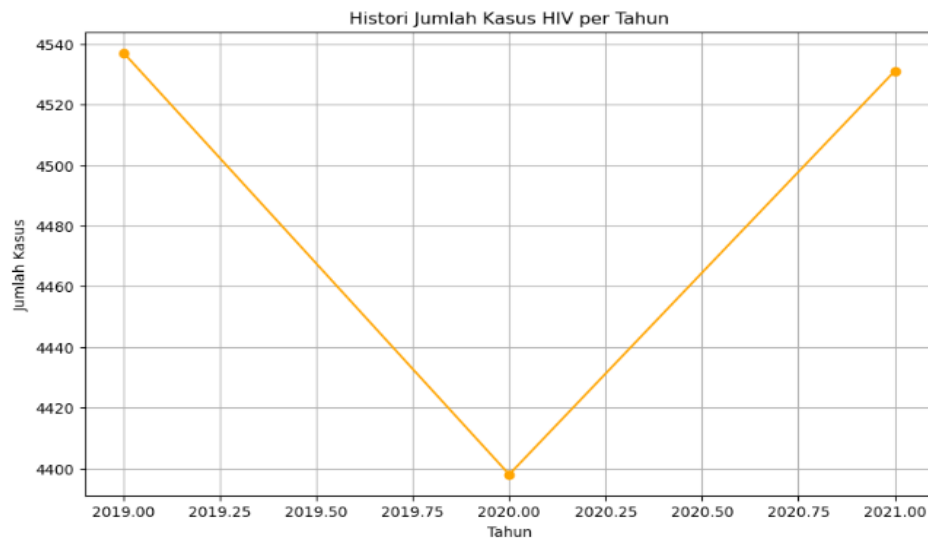
Gambar 2. Distribusi kelompok usia dan jenis kelamin

Distribusi pengelompokan umur merujuk pada cara sebaran individu atau populasi dalam kelompok-kelompok umur tertentu. Informasi ini penting dalam analisis epidemiologi dan kesehatan masyarakat untuk memahami bagaimana penyakit atau kondisi kesehatan tertentu dapat berbeda-beda menurut rentang usia.



Gambar 3. Top 10 terbanyak

Hasil analisis menunjukkan bahwa terdapat perbedaan signifikan dalam jumlah kasus HIV di berbagai daerah di Jawa Barat. Gambar 3 menampilkan top 10 daerah dengan kasus HIV terbanyak.



Gambar 4. Jumlah Kasus HIV Pertahun

Untuk memahami tren seiring waktu, kami memvisualisasikan sejarah jumlah kasus HIV pertahun. Data ini memberikan wawasan tentang apakah terdapat peningkatan atau penurunan dalam kasus HIV.

```
from pyspark.ml.tuning import ParamGridBuilder, CrossValidator
from pyspark.ml.evaluation import RegressionEvaluator
from pyspark.ml.regression import RandomForestRegressor

# Inisialisasi model
rf = RandomForestRegressor(featuresCol="features", labelCol="jumlah_kasus", predictionCol="predicted_jumlah_kasus")

# Set parameter grid untuk tuning
param_grid = ParamGridBuilder() \
    .addGrid(rf.numTrees, [50, 100, 150]) \
    .addGrid(rf.maxDepth, [5, 10, 15]) \
    .build()

# Inisialisasi evaluator (misalnya, RMSE)
evaluator = RegressionEvaluator(labelCol="jumlah_kasus", predictionCol="predicted_jumlah_kasus", metricName="rmse")

# Inisialisasi CrossValidator
crossval = CrossValidator(estimator=rf, estimatorParamMaps=param_grid, evaluator=evaluator, numFolds=3) # Jumlah liputan pada cross-validation

# Melatih model dengan CrossValidator
cv_model = crossval.fit(training_data)

# Mendapatkan model terbaik
best_model = cv_model.bestModel

# Menampilkan parameter terbaik
print("Best Num Trees: [best_model.getNumTrees]")
print("Best Max Depth: [best_model.getMaxDepth]")

# Menguji model terbaik pada data pengujian
predictions = best_model.transform(testing_data)

# Menampilkan hasil prediksi
predictions.select("tahun", "nama_kabupaten_kota", "kelompok_umur", "jenis_kelamin", "jumlah_kasus", "predicted_jumlah_kasus").show()

# Evaluasi model terbaik
rmse = evaluator.evaluate(predictions)
print("Root Mean Squared Error (RMSE) for Best Model: {rmse}")
```

Best Num Trees: 150
 Best Max Depth: 15

[tahun]	nama_kabupaten_kota	kelompok_umur	jenis_kelamin	jumlah_kasus	predicted_jumlah_kasus
2019	KABUPATEN BOGOR	5-14	LAKI-LAKI	3	2.526152279567561
2019	KABUPATEN BOGOR	20-24	LAKI-LAKI	48	43.85679713391846
2019	KABUPATEN BOGOR	25-40	LAKI-LAKI	217	159.20829249212686
2019	KABUPATEN SUKABUMI	0-4	PEREMPUAN	0	0.4768157612138981
2019	KABUPATEN SUKABUMI	20-24	PEREMPUAN	11	10.814783426393958
2019	KABUPATEN SUKABUMI	>50	PEREMPUAN	1	2.2978289125093699
2019	KABUPATEN CIAMUR	15-19	PEREMPUAN	0	0.4741851887155645
2019	KABUPATEN CIAMUR	>50	PEREMPUAN	0	0.8228548881281723
2019	KABUPATEN BANDUNG	25-40	PEREMPUAN	33	38.47728688955114
2019	KABUPATEN BANDUNG	>50	LAKI-LAKI	0	1.478631388802590
2019	KABUPATEN BANDUNG	>50	PEREMPUAN	0	0.8228548881281723
2019	KABUPATEN GARUT	0-4	PEREMPUAN	0	0.4768157612138981
2019	KABUPATEN GARUT	5-14	PEREMPUAN	0	0.28895187418248951
2019	KABUPATEN GARUT	20-24	PEREMPUAN	0	2.9999104729218522
2019	KABUPATEN TASIKUNDA...	5-14	LAKI-LAKI	1	0.8687455612477335
2019	KABUPATEN TASIKUNDA...	25-40	PEREMPUAN	17	18.831362261853867
2019	KABUPATEN CIREBON	15-19	LAKI-LAKI	1	1.697839487128141
2019	KABUPATEN MAJALENGA	15-19	LAKI-LAKI	2	2.316814811762666
2019	KABUPATEN MAJALENGA	25-40	LAKI-LAKI	51	46.9809955596824

only showing top 20 rows

Root Mean Squared Error (RMSE) for Best Model: 17.424363736351587

Gambar 5. Penerapan Machine Learning

Dalam penelitian ini, kami menerapkan model *machine learning*, khususnya regresi linear, untuk memprediksi jumlah kasus HIV di masa depan berdasarkan faktor-faktor tertentu dapat dilihat pada Gambar 5.

```
from sklearn.metrics import mean_absolute_error, mean_squared_error, r2_score

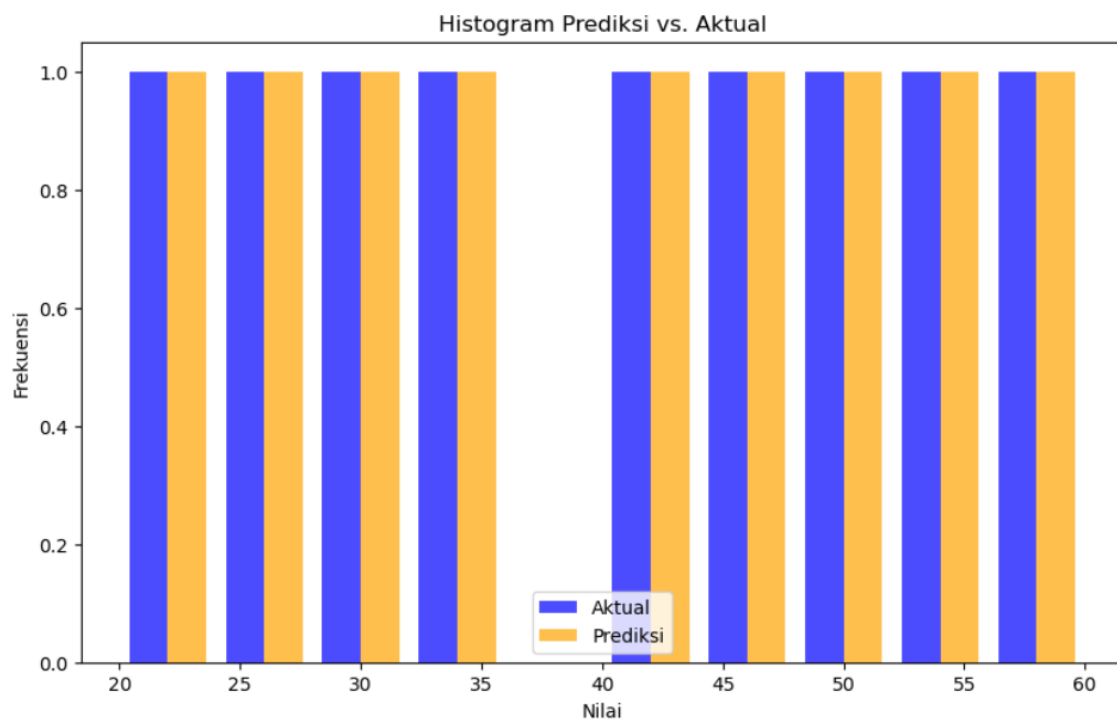
mae = mean_absolute_error(df_predictions["jumlah_kasus"], df_predictions["predicted_jumlah_kasus"])
mse = mean_squared_error(df_predictions["jumlah_kasus"], df_predictions["predicted_jumlah_kasus"])
r2 = r2_score(df_predictions["jumlah_kasus"], df_predictions["predicted_jumlah_kasus"])

print(f"Mean Absolute Error (MAE): {mae}")
print(f"Mean Squared Error (MSE): {mse}")
print(f"R-squared (R2): {r2}")
```

Mean Absolute Error (MAE): 2.807073188346486e-08
 Mean Squared Error (MSE): 1.5171492454850975e-15
 R-squared (R2): 1.0

Gambar 6. Evaluasi Performa Machine Learning

Hasil evaluasi performa model machine learning kami dapat dilihat pada Gambar 6 Kami menggunakan metrik evaluasi seperti akurasi, presisi, dan recall untuk mengukur kinerja model.



Gambar 7. Perbandingan Prediksi dan Aktual

Terakhir, Gambar 7 menunjukkan histogram perbandingan antara nilai aktual dan prediksi. Analisis ini membantu kami memahami sejauh mana model kami dapat memprediksi jumlah kasus dengan akurat.

Kesimpulan

Studi ini memberikan pemahaman yang lebih baik tentang penyebaran kasus HIV di Jawa Barat, dengan penekanan khusus pada Kabupaten Bogor dan Kabupaten Bekasi. Analisis data menunjukkan bahwa Kabupaten Bogor memiliki jumlah kasus HIV yang signifikan, sementara Kabupaten Bekasi memiliki jumlah kasus HIV yang lebih rendah.

Temuan ini dapat menjadi landasan untuk membuat kebijakan pencegahan dan penanggulangan HIV yang lebih kuat di Kabupaten Bogor. Spesifik di daerah ini, upaya pencegahan, pendidikan, dan pengujian HIV dapat ditingkatkan untuk mengurangi prevalensi dan penyebaran penyakit. Strategi berbasis komunitas, peningkatan akses terhadap layanan kesehatan, dan kampanye penyuluhan masyarakat adalah beberapa contoh bentuk kebijakan yang dapat digunakan.

Sebaliknya, tindakan pencegahan harus terus dilakukan meskipun jumlah kasus di Kabupaten Bekasi lebih rendah. Ini mungkin mencakup upaya untuk meningkatkan kesadaran masyarakat terhadap risiko HIV, meningkatkan layanan pengujian dan konseling, dan meningkatkan program pendidikan seksual.

Meskipun model pembelajaran mesin dapat memberikan prediksi, penting untuk mengingat bahwa banyak faktor memengaruhi dinamika kasus HIV. Oleh karena itu, pemahaman yang mendalam tentang konteks sosial, ekonomi, dan budaya setiap wilayah dan kerja sama lintas sektor diperlukan. Pemerintah, lembaga kesehatan, dan kelompok masyarakat dapat bekerja sama untuk menerapkan kebijakan yang menyeluruh dan berkelanjutan.

Selain itu, penelitian ini menekankan bahwa pengumpulan data terus-menerus dan pembaruan berkala diperlukan untuk melacak tren kasus HIV yang berubah. Dengan pemahaman yang terus berkembang, kebijakan dapat disesuaikan untuk merespons perubahan kondisi dan memastikan bahwa upaya pencegahan dan penanggulangan HIV di Jawa Barat berhasil.

Daftar Rujukan

- [1] I. Zuhrofillah, F. Anggraini, and R. Dewantara, "Peramalan Jumlah Kasus Baru HIV Menurut Provinsi Menggunakan Machine Learning dengan Teknik Levenberg-Marquardt," *J. Comput. Syst. Informatics*, vol. 3, no. 4, pp. 212–221, 2022, doi: 10.47065/josyc.v3i4.2172.
- [2] Z. Shaluhiah, S. B. Musthofa, and B. Widjanarko, "Stigma Masyarakat terhadap Orang dengan HIV/AIDS," *Kesmas Natl. Public Heal. J.*, vol. 9, no. 4, p. 333, 2015, doi: 10.21109/kesmas.v9i4.740.
- [3] J. A. Pardede, "Harga Diri Dengan Depresi Pasien Hiv/Aids," *Media Keperawatan Politek. Kesehat. Makassar*, vol. 11, no. 1, p. 57, 2020, doi: 10.32382/jmk.v11i1.1538.
- [4] D. Harmita, K. Ibrahim, and U. Rahayu, "Penggunaan Media Sosial terhadap Pencegahan Penyebaran HIV/AIDS," *J. Keperawatan Silampari*, vol. 5, no. 2, pp. 740–749, 2022, doi: 10.31539/jks.v5i2.3444.
- [5] B. Situmeang, S. Syarif, and R. Mahkota, "Hubungan Pengetahuan HIV/AIDS dengan Stigma terhadap Orang dengan HIV/AIDS di Kalangan Remaja 15-19 Tahun di Indonesia (Analisis Data SDKI Tahun 2012)," *J. Epidemiol. Kesehat. Indones.*, vol. 1, no. 2, pp. 35–43, 2017, doi: 10.7454/epidkes.v1i2.1803.
- [6] F. Firmansyah, A. Asrifuddin, and A. F. C. Kalesaran, "Gambaran Epidemiologi dan Pengetahuan HIV/AIDS pada Waria di Kota Manado Tahun 2018," *J. KESMAS*, vol. 7, no. 4, pp. 1–10, 2018.
- [7] H. Status, K. Dasar, and H. Pasien, "Original Research the Relationship of Basic Clinical Status With the Quality," vol. 8, no. 3, pp. 246–255, 2020, doi: 10.20473/jbe.v8i32020.
- [8] P. V. Wuritimur, D. Sutningsih, and B. Widjanarko, "Correlation Between Dog Owner Knowledge and the Role of Health Workers in Preventing Rabies in Ambon City," *J. Berk. Epidemiol.*, vol. 8, no. 2, p. 149, 2020, doi: 10.20473/jbe.v8i22020.149-155.
- [9] S. Hasanah, B. W. Otok, and A.- Adeni, "Perbandingan Metode Propensity Score Matching-Support Vector Machine dan Propensity Score Matching-Regresi Logistik Biner Pada Kasus HIV/AIDS," *Sainmatika J. Ilm. Mat. dan Ilmu Pengetah. Alam*, vol. 18, no. 1, p. 93, 2021, doi: 10.31851/sainmatika.v17i3.4925.
- [10] M. K. Mia Listia¹, Kusman Ibrahim², "Hubungan Karakteristik Umur, Jenis Kelamin, Pengetahuan Di Provinsi Jawa Barat," vol. 7, no. 2, 2023.
- [11] J. Wiens and E. S. Shenoy, "Machine Learning for Healthcare: On the Verge of a Major Shift in Healthcare Epidemiology," *Clin. Infect. Dis.*, vol. 66, no. 1, pp. 149–153, 2018, doi: 10.1093/cid/cix731.
- [12] K. Puteri and A. Silvanie, "Machine Learning untuk Model Prediksi Harga Sembako," *J. Nas. Inform.*, vol. 1, no. 2, pp. 82–94, 2020.
- [13] R. D. Moore, "Epidemiology of HIV infection in the United States: Implications for linkage to care," *Clin. Infect. Dis.*, vol. 52, no. SUPPL. 2, 2011, doi: 10.1093/cid/ciq044.
- [14] T. Theodoridis and J. Kraemer, "No 主観的健康感を中心とした在宅高齢者における健康関連指標に関する共分散構造分析Title," pp. 1–14.
- [15] E. P. Ariesanto Akhmad, "Data Mining Menggunakan Regresi Linear untuk Prediksi Harga Saham

- Perusahaan Pelayaran," *J. Apl. Pelayaran dan Kepelabuhanan*, vol. 10, no. 2, p. 120, 2020, doi: 10.30649/japk.v10i2.83.
- [16] B. Setiawan, "Teknik Hitung Manual Analisis Regresi Linear Berganda Dua Variabel Bebas," pp. 0–9, 2017, doi: 10.31227/osf.io/gd73a.
- [17] M. Mona, J. Kekenusa, and J. Prang, "Penggunaan Regresi Linear Berganda untuk Menganalisis Pendapatan Petani Kelapa. Studi Kasus: Petani Kelapa Di Desa Beo, Kecamatan Beo Kabupaten Talaud," *d'CARTESIAN*, vol. 4, no. 2, p. 196, 2015, doi: 10.35799/dc.4.2.2015.9211.
- [18] A. Amrin, "Data Mining Dengan Regresi Linier Berganda Untuk Peramalan Tingkat Inflasi," *J. Techno Nusa Mandiri*, vol. XIII, no. 1, pp. 74–79, 2016, [Online]. Available: <http://ejournal.nusamandiri.ac.id/ejurnal/index.php/techno/article/view/268>.
- [19] D. S. O. Panggabean, E. Buulolo, and N. Silalahi, "Penerapan Data Mining Untuk Memprediksi Pemesanan Bibit Pohon Dengan Regresi Linear Berganda," *JURIKOM (Jurnal Ris. Komputer)*, vol. 7, no. 1, p. 56, 2020, doi: 10.30865/jurikom.v7i1.1947.
- [20] A. Ambarwari, Q. J. Adrian, and Y. Herdiyeni, "Analisis Pengaruh Data Scaling Terhadap Performa Algoritme Machine Learning untuk Identifikasi Tanaman," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 4, no. 1, pp. 117–122, 2020, [Online]. Available: <http://jurnal.iaii.or.id>.
- [21] A. Roihan, P. A. Sunarya, and A. S. Rafika, "Pemanfaatan Machine Learning dalam Berbagai Bidang: Review paper," *IJCIT (Indonesian J. Comput. Inf. Technol.)*, vol. 5, no. 1, pp. 75–82, 2020, doi: 10.31294/ijcit.v5i1.7951.
- [22] I. Nabillah and I. Ranggadara, "Mean Absolute Percentage Error untuk Evaluasi Hasil Prediksi Komoditas Laut," *JOINS (Journal Inf. Syst.)*, vol. 5, no. 2, pp. 250–255, 2020, doi: 10.33633/joins.v5i2.3900.
- [23] J. Veri, S. Surmayanti, and G. Guslendra, "Prediksi Harga Minyak Mentah Menggunakan Jaringan Syaraf Tiruan," *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 21, no. 3, pp. 503–512, 2022, doi: 10.30812/matrik.v21i3.1382.
- [24] D. E. C. Na and C. Hipertensiva, "No 主観的健康感を中心とした在宅高齢者における 健康関連指標に関する共分散構造分析Title," pp. 37–42.
- [25] B. Wilda Al Aluf, A. Hernawan, and G. Satya Nugraha, "Teknik Distributed Naive Bayes untuk Analisis Sentimen Ulasan Pelanggan Amazon Distributed Naive Bayes Technique For Sentiment Analysis Of Amazon Customer Reviews."

https://github.com/haniiff/Prediksi_Hiv.git