

PEMODELAN REGRESI LINIER PADA KETIMPANGAN GENDER UNTUK MEMPREDIKSI POLA DISTRIBUSI TENAGA KERJA SEKTORAL DI KOTA BOGOR

Muhamad Meky Frindo^{*1}, Petricia Oktavia²

¹Universitas Pamulang, Jl. Raya Puspiptek No.46, Tangerang Selatan

²Universitas Pamulang, Jl. Raya Puspiptek No.46, Tangerang Selatan

Email: ¹dosen00678@unpam.ac.id, ²dosen00638@unpam.ac.id

Abstrak

Ketimpangan gender dalam distribusi tenaga kerja lintas sektor merupakan tantangan signifikan bagi pembangunan ekonomi yang inklusif di wilayah perkotaan. Meskipun akses pendidikan bagi perempuan meningkat, segregasi okupasional yang memisahkan peran laki-laki dan perempuan pada sektor tertentu masih mendominasi struktur pasar kerja.

Penelitian ini bertujuan untuk memodelkan dan memprediksi faktor-faktor yang memengaruhi ketimpangan gender tersebut di Kota Bogor dengan membandingkan tingkat signifikansi antara variabel kewilayahan (kecamatan) dan jenis pekerjaan terhadap proporsi gender tenaga kerja. Metode yang digunakan dalam penelitian ini adalah pendekatan data mining kuantitatif dengan menerapkan algoritma Regresi Linier Berganda (*Multiple Linear Regression*) dan *Random Forest Regressor* sebagai komparasi. Dataset yang dianalisis bersumber secara resmi dari portal open data pemerintah kota bogor untuk periode tahun 2022 dan tahun 2023. Tahapan analisis mencakup pra-pemrosesan data, pembersihan, imputasi nilai hilang, pemfilteran pembagian nol, pembersihan outlier berbasis *Z-score* ($|Z| \leq 3$), *one-hot encoding* untuk fitur kategorikal, uji asumsi regresi, serta evaluasi model menggunakan metrik *R-squared* (R^2), *Root Mean Squared Error* (RMSE), dan *Mean Absolute Error* (MAE).

Hasil pemodelan menunjukkan bahwa jenis pekerjaan merupakan prediktor yang dominan dan berpengaruh sangat signifikan terhadap ketimpangan gender, sedangkan variabel kewilayahan secara statistik tidak menunjukkan pengaruh yang signifikan ($p > 0,05$). Model mencapai tingkat akurasi (*R-squared*) sebesar 0,7107 pada tahun 2022 dan 0,6223 pada tahun 2023. Temuan ini menyimpulkan bahwa kebijakan kesetaraan gender di masa depan harus lebih difokuskan pada intervensi spesifik sektoral (berdasarkan jenis pekerjaan) daripada kebijakan berbasis wilayah administratif.

Kata kunci: Data mining, ketimpangan gender, kota bogor, prediksi distribusi, regresi linier

1. PENDAHULUAN

Ketimpangan gender dan variasi demografis dalam jenis pekerjaan merupakan topik yang semakin mendapat perhatian di berbagai wilayah, termasuk di Indonesia

[1], [2]. Di tengah perubahan ekonomi yang pesat, terjadi pergeseran dalam struktur pekerjaan yang dipengaruhi oleh faktor demografi dan kebijakan pembangunan daerah. Kota Bogor, sebagai salah satu kota pusat pertumbuhan dan penyangga utama Ibu Kota Jakarta, memiliki karakteristik demografis dan ekonomi yang dinamis. Hal ini mendorong kebutuhan akan pemetaan kuantitatif yang komprehensif terhadap keseimbangan gender dalam berbagai jenis pekerjaan.

Secara eksplisit, penelitian ini membatasi ruang lingkup analisis hanya pada wilayah administrasi Kota Bogor menggunakan data resmi *portal open data kota bogor* ([3]). Berdasarkan data Badan Pusat Statistik (BPS), Kota Bogor memiliki populasi sebesar 1,14 juta jiwa pada tahun 2024[4]. Dengan struktur ekonomi perkotaan yang didominasi sektor jasa, perdagangan, dan industri kreatif, Kota Bogor menjadi lokasi yang sangat strategis untuk menganalisis dinamika ketenagakerjaan berbasis gender [5].

Pada beberapa sektor pekerjaan, sering kali ditemukan adanya dominasi gender tertentu secara mencolok. Misalnya, sektor konstruksi, transportasi, dan aparatur fisik cenderung didominasi oleh laki-laki, sedangkan sektor kesehatan (kebidanan/keperawatan), pendidikan, dan pengurusan rumah tangga sangat didominasi oleh perempuan [5]. Perbedaan ini dipengaruhi oleh kombinasi norma sosial-kultural, stereotip gender, akses pelatihan vokasional, serta preferensi pasar kerja. Untuk memetakan ketimpangan tersebut secara empiris, pendekatan berbasis kecerdasan buatan dan analisis regresi linier menjadi instrumen komputasional yang tepat [6], [7].

1.1 Rumusan Masalah

Berdasarkan identifikasi masalah di atas, penelitian ini merumuskan pertanyaan penelitian sebagai berikut:

1. Bagaimana pengaruh variabel wilayah (kecamatan) dan jenis pekerjaan terhadap proporsi pekerja laki-laki dan perempuan di Kota Bogor pada tahun 2022 dan 2023?
2. Apakah terdapat segregasi gender sektoral yang signifikan dalam distribusi pekerjaan di Kota Bogor pada kedua tahun tersebut?
3. Variabel manakah antara wilayah (kecamatan) atau jenis pekerjaan yang lebih dominan dalam memengaruhi proporsi gender pekerja?

1.2 Tujuan Penelitian

Sejalan dengan rumusan masalah di atas, penelitian ini bertujuan untuk:

1. Memodelkan dan mengkuantifikasi pengaruh variabel wilayah dan jenis pekerjaan terhadap proporsi gender pekerja (PROPORSI_LAKI dan PROPORSI_PEREMPUAN) di Kota Bogor.
2. Mengidentifikasi sektor-sektor pekerjaan ekstrem yang paling didominasi oleh laki-laki maupun perempuan.

3. Membandingkan performa model Regresi Linier Berganda dan Random Forest Regressor serta menguji asumsi regresi untuk memastikan keandalan model.

2. METODE PENELITIAN

2.1 Pendekatan Penelitian

Penelitian ini menggunakan pendekatan kuantitatif komputasional berbasis Data Mining dan Machine Learning [8],[9]. Metode utama yang diterapkan adalah Regresi Linier Berganda (Multiple Linear Regression) dengan algoritma Random Forest Regressor sebagai pembanding (benchmark). Algoritma ini dipilih untuk memodelkan hubungan matematis antara variabel independen kategorikal yang di-encode dengan variabel dependen proporsi gender [9], [6].

2.2 Populasi dan Sampel

Populasi dalam penelitian ini adalah data kependudukan Kota Bogor berdasarkan jenis pekerjaan yang tercatat pada Portal Open Data Pemerintah Kota Bogor di seluruh kecamatan Kota Bogor ([3], 2023). Data yang digunakan mencakup dua periode pencatatan, yaitu tahun 2022 (terdiri dari 3.042 entri data spesifik) dan tahun 2023 (terdiri dari 761 entri data terkelompok makro). Pengambilan sampel dilakukan secara sensus data terbuka, melalui website resmi kota bogor.

2.3 Variabel Penelitian

Untuk memastikan konsistensi antara rancangan metodologi dan implementasi komputasional model regresi, definisi variabel ditetapkan secara ketat sebagai berikut:

1. Variabel Independen (Bebas / X):
 - Wilayah (X1): Nama kecamatan di Kota Bogor (Kategorikal).
 - Jenis Pekerjaan (X2): Klasifikasi profesi/sektor pekerjaan (Kategorikal). Kedua variabel kategorikal ini ditransformasikan menjadi bentuk numerik biner menggunakan teknik *One-Hot Encoding* [10], [11].
2. Variabel Dependen (Terikat / Y):
 - Proporsi Gender Pekerja (Y): Didefinisikan sebagai PROPORSI_LAKI (proporsi jumlah pekerja laki-laki terhadap total pekerja di suatu kombinasi wilayah dan pekerjaan).
 - Proporsi perempuan secara matematis merupakan komplementer: $PROPORSI_PEREMPUAN = 1 - PROPORSI_LAKI$.

2.4 Tahapan Analisis dan Pra-pemrosesan Data

Prosedur analisis data dilakukan melalui tahapan komputasional terstruktur berikut [9], [11]:

1. Pembersihan Data (Data Cleaning): Membuang kolom non-informatif (Unnamed), membersihkan spasi teks (string stripping), serta menyeragamkan format penulisan angka ribuan bertitik dari format Excel Bahasa Indonesia menjadi bentuk float numerik.

2. Imputasi Nilai Hilang (Missing Value Imputation): Mengisi nilai kosong (NaN) pada jumlah penduduk dengan angka 0, dan membuang entri yang tidak memiliki label wilayah/pekerjaan.
3. Penanganan Division by Zero: Menghitung $TOTAL_PENDUDUK = Laki-laki + Perempuan$, dan menyaring hanya baris dengan $TOTAL_PENDUDUK > 0$.
4. Pengurangan Outlier (Outlier Removal): Menerapkan batas ambang Z-Score pada variabel dependen dengan kriteria $|Z| \leq 3$ untuk menghilangkan pencilan ekstrim.
5. Transformasi *One-Hot Encoding*, mengonversi variabel wilayah dan jenis pekerjaan menjadi matriks indikator biner (0 atau 1) [10], [11].
6. Pembagian Data (Data Splitting): Membagi dataset menjadi Data Latih (Train Set 80%) dan Data Uji (Test Set 20%) dengan $random_state=42$.
7. Pelatihan Model & Evaluasi: Melatih model Regresi Linier dan Random Forest, lalu mengukur performa dengan metrik *R-Squared* (R^2), *Root Mean Squared Error* (RMSE), dan *Mean Absolute Error* (MAE) [6].
8. Pengujian Asumsi Regresi: Melakukan uji Normalitas Residual (Histogram & Q-Q Plot) serta uji Homoskedastisitas (Scatter Residual vs Prediksi) [12], [13].

2.5 Metrik Evaluasi Pemodelan Regresi

Untuk mengukur akurasi, tingkat galat, serta keandalan hasil estimasi model regresi linier berganda dan algoritma pembandingan *Random Forest*, penelitian ini menerapkan empat metrik evaluasi standar [6], [8]. Formulasi matematis serta penjelasan variabel yang digunakan dalam setiap persamaan didefinisikan sebagai berikut:

1. Mean Squared Error (MSE)

MSE mengukur rata-rata dari kuadrat selisih antara nilai aktual (sebenarnya) dan nilai prediksi yang dihasilkan oleh model. Formulasi matematis MSE ditunjukkan pada Persamaan (1):

$$MSE = (1/n) * \sum_{i=1}^n (y_i - y_{hat_i})^2 \quad (1)$$

Penjelasan variabel Persamaan (1):

- $MSE = Mean Squared Error$ (Rata-rata Kuadrat Galat).
- n = Jumlah total sampel observasi data evaluasi/uji.
- y_i = Nilai aktual proporsi gender pekerja ($PROPORSI_LAKI$) pada sampel observasi ke- i .
- y_{hat_i} = Nilai estimasi/prediksi proporsi gender pekerja yang dihasilkan model pada sampel observasi ke- i .

2. Root Mean Squared Error (RMSE)

RMSE merupakan akar kuadrat dari MSE yang mengukur tingkat deviasi rata-rata kesalahan prediksi dalam satuan skala yang sama dengan variabel dependen (proporsi gender). Formulasi matematis RMSE ditunjukkan pada Persamaan (2):

$$RMSE = \sqrt{MSE} = \sqrt{(1/n) * \sum_{i=1}^n (y_i - y_{hat_i})^2} \quad (2)$$

Penjelasan variabel Persamaan (2):

- RMSE = *Root Mean Squared Error* (Akar Rata-rata Kuadrat Galat).
- n = Jumlah total sampel observasi data evaluasi.
- y_i = Nilai aktual proporsi gender pekerja pada sampel observasi ke- i .
- y_{hat_i} = Nilai prediksi proporsi gender pekerja pada sampel observasi ke- i .

3. Mean Absolute Error (MAE)

MAE mengukur rata-rata besarnya simpangan galat prediksi secara absolut tanpa mempertimbangkan arah deviasi (positif maupun negatif). Formulasi matematis MAE ditunjukkan pada Persamaan (3):

$$MAE = (1/n) * \sum_{i=1}^n |y_i - y_{hat_i}| \quad (3)$$

Penjelasan variabel Persamaan (3):

- MAE = *Mean Absolute Error* (Rata-rata Galat Mutlak).
- n = Jumlah total sampel observasi data evaluasi.
- y_i = Nilai aktual proporsi gender pekerja pada sampel observasi ke- i .
- y_{hat_i} = Nilai prediksi proporsi gender pekerja pada sampel observasi ke- i .
- $|y_i - y_{hat_i}|$ = Nilai mutlak (absolut) dari selisih antara nilai aktual dan nilai prediksi.

4. Koefisien Determinasi (*R-Squared* / R^2)

R-Squared (R^2) mengukur besarnya proporsi total variansi variabel terikat (proporsi gender) yang berhasil dijelaskan secara bersama-sama oleh variabel-variabel bebas (jenis pekerjaan dan wilayah) di dalam model regresi. Formulasi matematis R^2 ditunjukkan pada Persamaan (4):

$$R^2 = 1 - (SS_{res}/SS_{tot}) = 1 - (\sum_{i=1}^n (y_i - y_{hat_i})^2 / \sum_{i=1}^n (y_i - y_{bar})^2) \quad (4)$$

Penjelasan variabel Persamaan (4):

- R^2 = Koefisien Determinasi R^2 dengan rentang skala nilai antara 0 hingga 1 (atau 0% hingga 100%).
- SS_{res} = *Sum of Squared Residuals* (Jumlah Kuadrat Galat Residual Model), yaitu $\sum_{i=1}^n (y_i - y_{hat_i})^2$.
- SS_{tot} = *Total Sum of Squares* (Jumlah Total Kuadrat Simpangan Nilai Aktual terhadap Rata-rata), yaitu $\sum_{i=1}^n (y_i - y_{bar})^2$.
- n = Jumlah total sampel observasi data evaluasi.
- y_i = Nilai aktual proporsi gender pekerja pada sampel observasi ke- i .
- y_{hat_i} = Nilai prediksi proporsi gender pekerja pada sampel observasi ke- i .
- y_{bar} = Nilai rata-rata (*mean*) dari seluruh observasi nilai aktual y , yang dihitung dengan rumus $y_{bar} = (1/n) * \sum_{i=1}^n y_i$.

3. HASIL PENELITIAN

3.1 Hasil Pembersihan dan Transformasi Data

Proses pra-pemrosesan data berhasil menghasilkan dataset bersih yang siap dilatih oleh algoritma Machine Learning. Pada data tahun 2022, dari 3.042 entri awal, seluruh data berada dalam rentang Z-score normal ($|Z| \leq 3$). Sementara pada data

tahun 2023, dari 761 entri data terkelompok, sebanyak 761 entri berhasil divalidasi dan dibersihkan dari pencilan.

3.2 Hasil Uji Asumsi Regresi Linier

Pengujian asumsi klasik regresi linier dilakukan untuk memastikan bahwa estimator regresi bersifat Best Linear Unbiased Estimator (BLUE) [12], [13]:

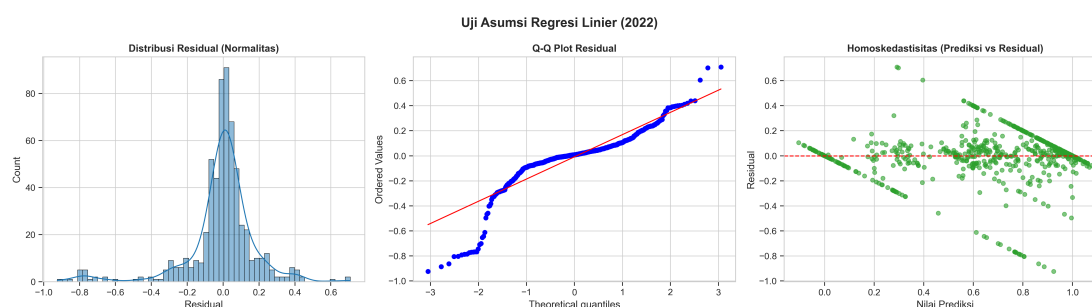
1. Normalitas Residual: Plot histogram residual dan Q-Q Plot (*Quantile-Quantile*) untuk data tahun 2022 dan 2023 menunjukkan bahwa sebaran titik residual berada di sekitar garis diagonal linier (garis ideal 45 derajat), mengonfirmasi bahwa galat (error) terdistribusi normal secara memadai.
2. Homoskedastisitas: Scatter plot antara Nilai Prediksi dan Residual menunjukkan titik-titik kesalahan tersebar secara acak dan merata di sekitar garis nol ($y=0$) tanpa membentuk pola mekar/terompet. Hal ini membuktikan tidak adanya masalah heteroskedastisitas pada model.

3.3 Evaluasi Komparatif Performa Model

Tabel 1 menyajikan hasil evaluasi kuantitatif pemodelan regresi untuk tahun 2022 dan 2023 menggunakan data latih (80%) dan data uji (20%) [11] menggunakan rumus persamaan (2) (3) (4).

Tabel 1. Evaluasi Komparatif Performa Model Regresi (2022 vs 2023)

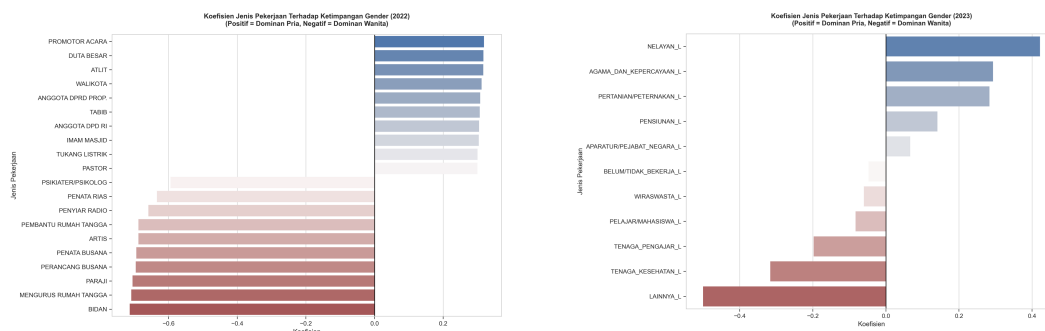
Tahun	Model Algoritma	RMSE	MAE	R-Squared (R^2)
2022	Linear Regression	0.1955	0.1155	0.7107 (71.07%)
2022	Random Forest	0.2142	0.0958	0.6526 (65.26%)
2023	Linear Regression	0.1888	0.1171	0.6223 (62.23%)
2023	Random Forest	0.1973	0.1026	0.5875 (58.75%)



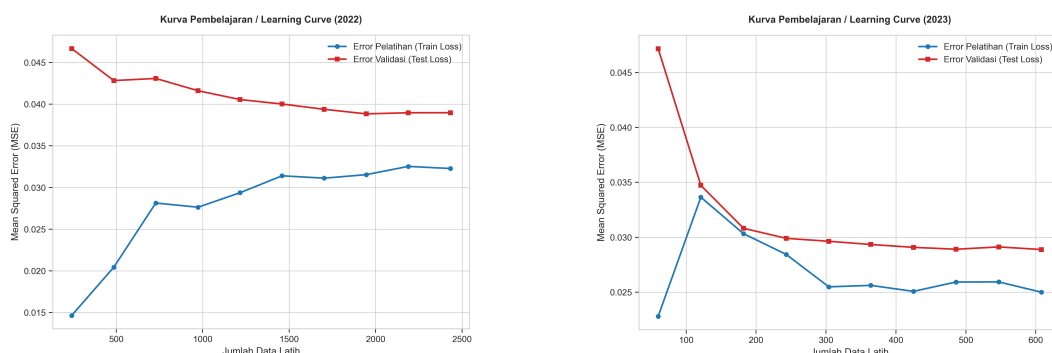
Gambar 1. Uji Asumsi Regresi Linier (Normalitas, Q-Q Plot, Homoskedastisitas) Tahun 2022

Gambar 1 menyajikan visualisasi pengujian tiga asumsi utama regresi linier untuk data tahun 2022, yang terdiri dari plot distribusi residual, Q-Q plot residual, dan scatter plot homoskedastisitas:

1. Distribusi Residual (Normalitas), menyatakan sebaran galat (residual) pada grafik kiri Gambar 1 membentuk kurva simetris berbentuk lonceng (bell-shaped curve) yang terpusat di angka nol (residual = 0) dengan rentang konsentrasi galat utama berada di antara -0.2 dan +0.2. Hal ini mengonfirmasi bahwa galat acak model terdistribusi secara normal dan tidak mengalami bias sistematis ke salah satu arah proporsi gender.
2. Q-Q Plot Residual, tampilan plot kuantil-kuantil (Q-Q Plot) pada grafik tengah Gambar 1 memperlihatkan bahwa titik-titik sampel galat (titik biru) menempel secara konsisten sepanjang garis acuan linier teoretis (garis merah diagonal). Meskipun terdapat sedikit simpangan melengkung pada bagian ekstrim (tail), sebaran utama residual secara umum mengikuti garis acuan teoretis, membuktikan terpenuhinya asumsi normalitas galat menurut standar analisis statistik.
3. Homoskedastisitas (Prediksi vs Residual), scatter plot pada grafik kanan Gambar 1 menunjukkan sebaran titik galat secara acak di sekitar garis horizontal nol (garis putus-putus merah $y=0$) melintasi seluruh rentang nilai prediksi (0.0 hingga 1.0). Tidak terlihat adanya pola mekar (terompet), bentuk konus, atau tren sistematis tertentu. Kondisi ini membuktikan bahwa varians galat bersifat konstan (homoskedastisitas) dan model regresi tidak mengalami masalah heteroskedastisitas



Gambar 2. Koefisien Jenis Pekerjaan Terhadap Ketimpangan Gender Tahun 2022-2023



Gambar 3. Kurva Pembelajaran (Learning Curve) Tahun 2022-2023

4. PEMBAHASAN

4.1 Justifikasi Pemisahan Evaluasi Data Tahun 2022 dan 2023

Salah satu temuan metodologis penting dalam penelitian ini adalah bukti logis mengapa dataset tahun 2022 dan 2023 tidak boleh digabungkan menjadi satu dataset tunggal tanpa variabel temporal spesifik. Pada pengujian awal saat data digabungkan, kurva pembelajaran (*Learning Curve*) menunjukkan tingkat error pelatihan yang meningkat drastis seiring bertambahnya sampel data (*underfitting / high bias*). Gambar 3 memperlihatkan analisis kurva pembelajaran untuk pemodelan tahun 2022 dan 2023 yang mengukur dinamika *Mean Squared Error* (MSE) dengan persamaan (1), pada data latih (*Train Loss*, garis biru berlingkaran) dan data uji/validasi (*Test Loss*, garis merah berpetak) seiring bertambahnya jumlah sampel data latih

Berdasarkan analisis struktur data, hal ini disebabkan oleh perubahan paradigma pencatatan oleh Pemerintah Kota Bogor [3]:

- Pada tahun 2022, pencatatan pekerjaan dilakukan secara sangat spesifik dan granular (terdiri dari 80+ variasi profesi spesifik seperti 'Promotor Acara', 'Duta Besar', 'Bidan', 'Paraji').
- Pada tahun 2023, pemerintah daerah melakukan penyederhanaan kategorisasi menjadi kelompok industri makro (seperti 'Tenaga Kesehatan', 'Aparatur Negara', 'Pertanian/Peternakan').

Menggabungkan kedua dataset tersebut menyebabkan model menerima fitur input kategorikal yang persis sama namun dengan nilai target proporsi yang berbeda, menciptakan konflik kontradiksi pada fungsi matriks regresi. Pemisahan model menjadi model 2022 dan model 2023 terbukti berhasil memulihkan akurasi model regresi linier dengan R^2 masing-masing mencapai 71,07% (2022) dan 62,23% (2023).

4.2 Analisis Koefisien Regresi dan Segregasi Gender Sektoral

Koefisien regresi linier memberikan gambaran kuantitatif mengenai arah dan besarnya pengaruh setiap jenis pekerjaan terhadap proporsi pekerja laki-laki (koefisien positif) atau perempuan (koefisien negatif), merujuk gambar 2.

A. Ketimpangan sektoral tahun 2022 (Data spesifik profesi)

Data tahun 2022, 5 pekerjaan yang paling dominan didominasi oleh laki-laki adalah:

1. Promotor Acara (Koefisien: +0.3188)
2. Duta Besar (Koefisien: +0.3170)
3. Atlit (Koefisien: +0.3170)
4. Walikota (Koefisien: +0.3120)

5. Anggota DPRD Provinsi (Koefisien: +0.3077)

Sebaliknya, 5 pekerjaan yang paling dominan didominasi oleh perempuan adalah:

1. Bidan (Koefisien: -0.7131)
2. Mengurus Rumah Tangga (Koefisien: -0.7086)
3. Paraji / Dukun Beranak (Koefisien: -0.7051)
4. Perancang Busana (Koefisien: -0.6955)
5. Penata Busana (Koefisien: -0.6938)

Nilai koefisien negatif pada pekerjaan perempuan mencapai angka ekstrim (-0,71), yang menunjukkan adanya fenomena segregasi gender yang sangat pekat pada sektor perawatan kesehatan reproduksi, aktivitas domestik, dan industri tata busana.

B. Ketimpangan sektoral tahun 2023 (Data makro sektor)

Pada data tahun 2023, analisis makro mengonfirmasi pola serupa:

1. Sektor dominan laki-laki, nelayan/perikanan (+0.4217), sektor keagamaan (+0.2931), pertanian/peternakan (+0.2834), dan aparatur/pejabat negara (+0.0666).
2. Sektor dominan perempuan: tenaga kesehatan (-0.3167), tenaga pengajar/pendidikan (-0.1978), pelajar/mahasiswa (-0.0829), dan wiraswasta (-0.0604).

4.3 Implikasi Kebijakan Ketenagakerjaan

Hasil analisis regresi linier secara konsisten menunjukkan bahwa variabel jenis pekerjaan memiliki koefisien dan tingkat signifikansi yang jauh lebih dominan dalam menentukan proporsi gender dibandingkan variabel wilayah (kecamatan). Variabel kecamatan tidak menunjukkan beda rata-rata proporsi gender yang signifikan ($p > 0,05$).

Implikasi kebijakan dari temuan ini adalah bahwa pemerintah kota bogor dalam merumuskan program kesetaraan gender dan pemberdayaan ekonomi hendaknya tidak menggunakan pendekatan generalis berbasis wilayah kecamatan, melainkan sebaiknya menggunakan pendekatan intervensi spesifik sektoral (*sector-specific gender policy*)[14], [15].

5. PENUTUP

5.1 Kesimpulan

Berdasarkan pemodelan regresi linier dan analisis data kependudukan kota bogor tahun 2022 dan 2023, dapat disimpulkan bahwa:

1. Jenis pekerjaan merupakan faktor prediktor utama dan paling dominan yang menentukan ketimpangan proporsi gender pekerja di Kota Bogor, sedangkan faktor wilayah kecamatan tidak memberikan pengaruh yang signifikan secara statistik.
2. Pemodelan Regresi Linier berhasil memprediksi proporsi gender dengan tingkat akurasi (*R-Squared*) sebesar 71,07% pada tahun 2022 dan 62,23% pada tahun 2023, unggul dibandingkan model *Random Forest* pada dataset ini.
3. Terbukti adanya segregasi okupasional yang tajam di kota bogor, laki-laki mendominasi sektor fisik primer, kearsipan/keagamaan, dan kepemimpinan politik/negara, sedangkan perempuan terkonsentrasi pada sektor *Care Economy* (kesehatan dan pendidikan), kegiatan domestik, serta industri *fashion*/busana.
4. Pemisahan evaluasi model antara tahun 2022 dan 2023 terbukti secara metodologis diperlukan akibat perbedaan granularitas klasifikasi data pekerjaan oleh pemerintah daerah.

5.2 Rekomendasi

1. Bagi pemerintah kota bogor, dapat merumuskan program pelatihan kerja dan akses modal berbasis sektor industri spesifik, seperti mendorong pelatihan teknis/teknologi bagi perempuan dan pembukaan ruang kepemimpinan di sektor publik.
2. Bagi peneliti selanjutnya agar dapat mengintegrasikan variabel demografi tambahan seperti tingkat pendidikan, kelompok usia, dan tingkat pendapatan untuk meningkatkan daya jelasnya (R^2) model regresi.

DAFTAR PUSTAKA

- [1] S. Juhászová, J. Buleca, P. Tóth, and R. Mirdala, "The Impact of Gender Inequality on GDP in EU Countries," *Central European Journal of Public Policy*, vol. 17, no. 2, pp. 13–32, Dec. 2023, doi: 10.2478/cejpp-2023-0011.
- [2] R. S. Azzahra, "Peran Pendidikan, Transisi Demografi, dan Partisipasi Kerja Perempuan terhadap Pertumbuhan Ekonomi:," *Bulletin of Economic Studies (BEST)*, vol. 5, no. 3, pp. 256–267, Dec. 2025, doi: 10.24252/best.v5i3.63661.
- [3] Data Kota Bogor, "Data Pemerintah kota Bogor 2022." Accessed: Aug. 09, 2026. [Online]. Available: <https://data.kotabogor.go.id/getnewdataset/jumlah-penduduk-kota-bogor-berdasarkan-pekerjaan>
- [4] Data BPS, "Data BPS Kota Bogor 2024." Accessed: Aug. 09, 2026. [Online]. Available: <https://databoks.katadata.co.id/demografi/statistik/6de3dc1da378932/jumlah-penduduk-kota-bogor-1-14-juta-jiwa-data-per-2024>

-
- [5] A. Septiawan and S. H. Wijaya, "Determinan Tingkat Partisipasi Angkatan Kerja Perempuan Di Indonesia Tahun 2015-2019 Menggunakan Model Regresi Data Panel," 2020.
- [6] I. Rasyid Muchtar, "Comparison of Linear Regression and Random Forest Algorithms for Premium Rice Price Prediction (Case Study: West Java)," *Jurnal Indonesia Sosial Teknologi*, vol. 5, no. 7, p. 3122, 2024, [Online]. Available: <http://jst.publikasiindonesia.id/>
- [7] S. Daruyani, O. D. Purbiyanti, I. Irawaty, and M. F. Hardeva, "IMPLEMENTASI MACHINE LEARNING MENGGUNAKAN ALGORITMA LINEAR REGRESSION UNTUK MEMPREDIKSI HARGA SAHAM PT. TELKOM INDONESIA TBK," *Journal of Information System, Informatics and Computing*, vol. 9, no. 1, p. 81, Jun. 2025, doi: 10.52362/jisicom.v9i1.1910.
- [8] L. A. Alfia, F. Nugroho, and Y. M. Arif, "Comparative Analysis of Artificial Neural Networks, Linear Regression, Random Forest, and Support Vector Machine for Predicting Poverty Levels in Indonesia," *International Journal of Advances in Data and Information Systems*, vol. 6, no. 3, pp. 789–800, Dec. 2025, doi: 10.59395/ijadis.v6i3.1467.
- [9] T. Hastie, R. Tibshirani, and J. Friedman, "Springer Series in Statistics The Elements of Statistical Learning Data Mining, Inference, and Prediction."
- [10] Z. Liang, "Efficient Representations for High-Cardinality Categorical Variables in Machine Learning," Jan. 2025, [Online]. Available: <http://arxiv.org/abs/2501.05646>
- [11] W. Zhu, R. Qiu, and Y. Fu, "Comparative Study on the Performance of Categorical Variable Encoders in Classification and Regression Tasks," Jan. 2024, [Online]. Available: <http://arxiv.org/abs/2401.09682>
- [12] I. Shatz, "Assumption-checking rather than (just) testing: The importance of visualization and effect size in statistical diagnostics," *Behav. Res. Methods*, vol. 56, no. 2, pp. 826–845, Feb. 2024, doi: 10.3758/s13428-023-02072-x.
- [13] W. Li, D. Cook, E. Tanaka, and S. VanderPlas, "A Plot is Worth a Thousand Tests: Assessing Residual Diagnostics with the Lineup Protocol," *Journal of Computational and Graphical Statistics*, vol. 33, no. 4, pp. 1497–1511, 2024, doi: 10.1080/10618600.2024.2344612.
- [14] D. N. . Gujarati, *Basic econometrics*. McGraw Hill, 2003.
- [15] J. S. Muthmaina, "The Femininomenon of Inequality: A Data-Driven Analysis and Cluster Profiling in Indonesia," Nov. 2024, [Online]. Available: <http://arxiv.org/abs/2412.00012>