IMPLEMENTASI *DATA MINING* MENGGUNAKAN METODE REGRESI DATA PANEL UNTUK MEMPREDIKSI CAPAIAN INDEKS PEMBANGUNAN MANUSIA (STUDI KASUS: PROVINSI JAWA BARAT)

Salsa Khaerunisa, Tesa Nur Padilah, Jajam Haerul Jaman
Program Studi Informatika S1, Fakultas Ilmu Komputer
Universitas Singaperbangsa Karawang, Jalan HS. Ronggowaluyo, Karawang, Indonesia
1910631170133@student.unsika.ac.id

ABSTRAK

Pada tahun 2022, capaian IPM di 19 dari 27 kabupaten/kota Provinsi Jawa Barat tidak sesuai dengan target yang ditetapkan dalam Rencana Pembangunan Jangka Menengah Daerah (RPJMD). Fakta tersebut menunjukkan bahwa hanya 30% kabupaten/kota di Provinsi Jawa Barat yang berhasil mencapai target IPM yang diamanatkan, sementara 70% wilayah lainnya masih belum mencapai sasaran. Dari permasalahan tersebut, maka perlu dilakukan prediksi capaian IPM agar diketahui nilai prediksi IPM di masa mendatang sehingga pemerintah memiliki referensi dan acuan dalam membuat langkah-langkah strategis ataupun berbagai kebijakan terkait peningkatan IPM kabupaten/kota di Provinsi Jawa Barat. Penelitian ini menggunakan metode regresi data panel untuk memprediksi capaian IPM kabupaten/kota di Provinsi Jawa Barat tahun 2023. Data yang digunakan terdiri dari data capaian IPM, usia harapan hidup, harapan lama sekolah, rata-rata lama sekolah, dan pendapatan perkapita disesuaikan tahun 2010-2022 dari 27 kabupaten dan kota yang ada di Provinsi Jawa Barat. Hasil prediksi menunjukkan bahwa pada tahun 2023, 11 kabupaten/kota diprediksi mampu mencapai target IPM dan 16 kabupaten/kota lainnya diprediksi belum mampu mencapai target IPM sesuai dengan target RPJMD. Adapun hasil evaluasi menunjukkan model prediksi menggunakan metode regresi data panel layak digunakan untuk memprediksi capaian IPM dengan nilai MAPE sebesar 0,87%, MAD sebesar 0,56, dan MSD sebesar 0,76.

Kata kunci: *Data Mining, IPM, Prediksi, Regresi Data Panel*

1. PENDAHULUAN

Menurut *United Nation Development Programme* (UNDP), Indeks Pembangunan Manusia (IPM) digunakan untuk mengukur seberapa berhasil pembangunan manusia telah dicapai pada suatu wilayah. Secara teoritis, IPM adalah suatu pencapaian yang menggambarkan bagaimana individu memanfaatkan hasil pembangunan dalam dalam aspek pendapatan, kesehatan, pendidikan, dan variabel lainnya. IPM digunakan oleh pemerintah sebagai alat ukur kinerja dan sebagai faktor penentu dalam alokasi Dana Alokasi Umum (DAU).

Pemerintah Provinsi Jawa Barat menyatakan bahwa pedoman kebijakan pembangunan masing-masing kabupaten/kota dapat dicapai melalui lima indikator kinerja makro pemerintah, salah satunya adalah IPM. Berdasarkan data BPS, IPM Provinsi Jawa Barat terus meningkat setiap tahunnya. Persoalannya, meskipun IPM di Provinsi Jawa Barat terus meningkat setiap tahunnya, sasaran IPM tidak tercapai di semua kabupaten dan kota. Pada tahun 2022, capaian IPM di 19 dari total 27 kabupaten dan kota di Provinsi Jawa Barat tidak sesuai dengan target yang telah diamanatkan dalam Rencana Pembangunan Jangka Menengah Daerah (RPJMD). Fakta ini menunjukkan bahwa hanya 30% dari wilayah kabupaten dan kota di Provinsi Jawa Barat yang berhasil mencapai target IPM yang ditetapkan, sementara 70% wilayah lainnya masih belum mencapai sasaran tersebut.

Dari permasalahan yang ada, dimana IPM memiliki peran yang penting dalam ukuran kinerja pemerintah dan masih banyaknya kabupaten/kota Provinsi Jawa Barat yang belum mampu mencapai target IPM, maka perlu dilakukan prediksi capaian IPM Provinsi Jawa Barat agar diketahui nilai prediksi IPM di masa mendatang sehingga pemerintah memiliki referensi dan acuan dalam membuat langkah-langkah strategis ataupun mengambil berbagai kebijakan seperti kebijakan dalam penyaluran dana, penyediaan fasilitas, serta kebijakan terkait peningkatan IPM kabupaten/kota Provinsi Jawa Barat dalam periode pemerintahan selanjutnya.

Berdasarkan permasalahan yang telah dipaparkan diatas, penelitian ini akan mengimplementasikan data mining menggunakan metode regresi data panel untuk memprediksi capaian IPM kabupaten/kota di Provinsi Jawa Barat pada tahun 2023. Selain itu, agar informasi prediksi dapat dikomunikasikan secara jelas dan efisien dibuat juga visualisasi hasil prediksi melalui pemetaan capaian IPM kabupaten/kota Provinsi Jawa Barat menggunakan Power BI. Dengan adanya pemetaan tersebut, diharapkan dapat membantu pemerintah dan pihak-pihak yang bersangkutan dalam mengawasi perkembangan IPM di setiap kabupaten/kota dengan lebih efektif.

2. TINJAUAN PUSTAKA

2.1. *Data Mining*

Data mining merupakan proses penemuan pengetahuan dalam database menggunakan metodologi atau prosedur tertentu melalui pencarian pola pada data yang dipilih. *Data mining* merupakan suatu proses mencari dan mengekstraksi informasi berharga dari *database* berkapasitas besar dengan menggunakan metode statistik, matematika, kecerdasan buatan, dan pembelajaran mesin [1].

Menurut [2] berikut adalah beberapa metode *data mining* berdasarkan fungsinya, yaitu:

1. Asosiasi (*Association*).
   Digunakan untuk menemukan korelasi antara berbagai peristiwa atau proses untuk mendeteksi pola.
2. Pengklusteran (*Clustering*)
   Clustering adalah proses menyatukan data, pengamatan, atau kelas tergantung pada seberapa mirip data-data tersebut. *Cluster* adalah kumpulan *record* yang berbeda dari *record* di *cluster* lain namun dapat dibandingkan satu sama lain.
3. Prediksi (*Prediction*)
   Hampir mirip dengan klasifikasi dan estimasi, hanya saja prediksi mengantisipasi nilai yang akan diperoleh di masa depan.
4. Estimasi (*Estimation*)
   Perbedaan antara estimasi dan klasifikasi adalah bahwa variabel objektif untuk estimasi seringkali bersifat numerik dan bukan kategori. Nilai variabel target digunakan sebagai prediktor dalam model estimasi yang dibangun menggunakan dataset lengkap. Estimasi dibentuk berdasarkan nilai variabel prediksi.
5. Klasifikasi (*Classification*)
   Klasifikasi merupakan fungsi *data mining* yang menempatkan potongan data (item) pada salah satu kelas dari banyaknya kelas yang telah ditentukan sebelumnya.

2.2. *Knowledge Discovery in Database* (KDD)

Menurut [3] KDD merupakan prosedur *non-trivial* yang bermanfaat untuk mengidentifikasi pola dalam data ketika pola tersebut baru, valid, praktis, dan mudah dipahami. Informasi yang diperoleh melalui *data mining* dapat dimanfaatkan sebagai dasar pengambilan keputusan. KDD terdiri dari beberapa proses, yaitu:

1) Data Selection
Memilih data yang akan digunakan sebelum proses ekstraksi data menggunakan KDD.

2) Pre-processing (Cleaning)
Sebelum memulai proses penambangan data, data harus dibersihkan terlebih dahulu, yang mencakup penghapusan duplikasi data, mengecek apakah terdapat data yang dianggap tidak konsisten, serta mendeteksi dan memperbaiki kesalahan.

3) Transformation
Tahap transformasi merupakan proses mentransformasikan data yang dipilih agar data tersebut layak digunakan dalam proses *data mining* dengan mentransformasikan data tersebut menjadi model analisis dan membangun model sesuai dengan analisis yang diharapkan dan dibutuhkan dalam algoritma *data mining*.

4) Data Mining
Tahap *data mining* atau penambangan data adalah proses eksplorasi data melalui penemuan pola dan informasi unik pada sekumpulan data yang telah diseleksi dengan menggunakan metode atau teknik khusus.

5) Interpretation/Evaluation/Knowledge
Tahap pemeriksaan model atau data yang ditemukan untuk menentukan apakah bertentangan dengan hipotesis atau fakta yang ada.

2.3. Regresi Data Panel

Regresi data panel adalah analisis regresi linier berganda yang diterapkan pada data panel (kombinasi data *cross-sectional* dan data *time-series* dengan unit *cross-sectional* yang sama diukur pada banyak periode). Sehingga secara fungsional, pemanfaatan regresi data panel sama dengan regresi linier berganda, yaitu untuk memeriksa hubungan antara satu variabel dependen dengan satu atau lebih variabel independent [4].

Menurut [4] secara umum persamaan regresi data panel dituliskan sebagai berikut:

$$Y_{it} = \alpha_{it} + \beta X_{it} + e_{it} \quad (1)$$

Keterangan:

Y_{it} : mewakili unit cross section ke-i untuk periode waktu ke-t".

α_{it} : adalah intersep yang mewakili efek grup atau individu dari unit cross section ke-i dan periode waktu ke-t

$\beta = (\beta 1, \beta 2, \dots, \beta n)$ adalah vektor konstanta berukuran $1 \times n$ dimana n merupakan banyaknya variabel independen

X_{it} : mengindikasikan vektor observasi pada variabel independen dengan dimensi 1 x n

e_{it} : komponen eror unit data silang ke-i dan waktu ke-t

$i = 1,2,\dots,n$

$t = 1.2,...,T$

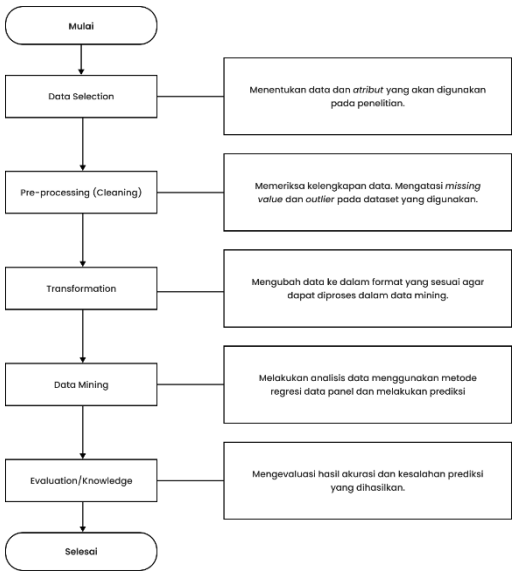
2.4. Power BI

Power BI adalah sebuah alat *business intelegent* yang memiliki fokus utama pada visualisasi dan eksplorasi data. Power BI mendukung metodologi pemodelan data, seperti transformasi data, pembersihan data, pembentukan data, dan lain sebagainya. Selain itu, Power BI juga mendukung metodologi analitik tingkat lanjut seperti pembelajaran mesin dan analitik prediktif [5].

Beberapa visualisasi data yang tersedia antara lain grafik berbaris, grafik batang, pie chart, donut

chart, peta, dan scatter plot. Dengan tampilan antarmuka Power BI yang mudah digunakan, pengguna dapat dengan mudah berkolaborasi dalam membuat laporan interaktif, memvisualisasikan data, dan berbagi temuan [6].

3. METODE PENELITIAN



Gambar 1. Metode Penelitian

Penelitian ini dibangun berdasarkan metode *Knowledge Discovery in Database* (KDD) yang diawali dengan tahap *data selection*, *pre-processing*, *transformation*, *data mining*, dan diakhiri dengan *evaluation/knowledge.*

3.1. *Data Selection*

Pada tahap ini dilakukan proses pemilihan data yang relevan dengan penelitian, yaitu dengan mengunjungi *website* Badan Pusat Statistik (BPS) Provinsi Jawa Barat.

3.2. *Pre-processing/Data Cleaning*

Setelah memilih data yang akan digunakan, langkah selanjutnya yaitu melakukan pemeriksaan *missing value* pada data penelitian. Apabila selama proses pengecekan ditemukan *missing value*, maka metode *Multiple Imputation by Chained Equation* (MICE) digunakan sebagai solusi untuk mengatasi *missing value.* Disamping itu, dilakukan juga pemeriksaan *outlier* pada tiap variabel penelitian menggunakan *boxplot* sehingga jumlah dan nilai-nilai data yang merupakan *outlier* dapat terdeteksi untuk kemudian dibersihkan menggunakan metode *winsorizing* dengan memanfaatkan *Inter Quartile Range* (IQR).

3.3. *Transformation*

Pada tahap ini, dilakukan proses tranformasi pada data yang telah melalui tahap *selection* dan *cleaning* agar data tersebut layak digunakan pada proses *data mining*. Dikarenakan metode yang digunakan pada penelitian ini merupakan regresi data panel, maka tahap ini dilakukan dengan memastikan dan melakukan tranformasi agar tipe data tiap variabel merupakan numerik.

3.4. *Data Mining*

Tahap ini adalah tahap analisis data, yang meliputi serangkaian langkah sebagai berikut:

1) Pemilihan model data panel dengan melakukan uji *Chow*, uji *Hausman*, dan uji *Lagrange Muliplier*.
2) Melakukan uji asumsi klasik yang meliputi uji normalitas, uji multikolinieritas, uji heteroskedastisitas dan uji autokorelasi.
3) Melakukan uji F untuk menguji signifikansi secara simultan, serta uji T untuk menguji signifikansi secara parsial.
4) Menghitung koefisien determinasi untuk menilai sejauh mana keterkaitan antara variabel independen dengan variabel dependen.
5) Menentukan nilai variabel independen masing-masing kabupaten dan kota pada tahun 2023 menggunakan analisis tren linier, kuadratik, dan eksponensial.
6) Menghitung nilai prediksi menggunakan persamaan regresi dan nilai variabel independen yang ada.

3.5. *Evaluation/Knoeledge*

Evaluasi dilakukan untuk mengetahui nilai eror yang dihasilkan dari model prediksi sehingga dapat diketahui tingkat akurasi prediksi capaian IPM kabupaten dan kota Provinsi Jawa Barat tahun 2023 menggunakan metode regresi data panel. Evaluasi model prediksi dilakukan menggunakan *Mean Absolute Percentage Error* (MAPE), *Mean Absolute Deviation* (MAD), dan *Mean Squared Deviation* (MSD). Hasil prediksi dan evaluasi kemudian diinterpretasikan supaya dapat dimengerti oleh para pihak yang memiliki kepentingan melalui pembuatan peta capaian IPM kabupaten/kota Provinsi Jawa Barat tahun 2023 menggunakan Power BI.

4. HASIL DAN PEMBAHASAN

4.1. *Data Selection*

Data yang digunakan pada penelitian ini adalah data sekunder dari Badan Pusat Statistik Provinsi Jawa Barat, yaitu data capaian IPM, usia harapan hidup, harapan lama sekolah, rata-rata lama sekolah, dan pendapatan perkapita disesuaikan tahun 2010-2022. Tabel 1 berisi atribut yang digunakan pada penelitian ini.

Adapun variabel dependen yang digunakan pada penelitian ini adalah capaian IPM kabupaten/kota Provinsi Jawa Barat. Sedangkan variabel independen yang digunakan pada penelitian adalah usia harapan hidup, harapan lama sekolah, rata-rata lama sekolah, dan pendapatan perkapita disesuaikan tahun 2010-2022.

Tabel 1. Atribut penelitian

Nama Atribut	Tipe Data	Keterangan
Tahun	Numerik	Tahun perolehan data
Kabupaten/kota	Nominal	Nama kabupaten/kota
UHH	Numerik	Usia Harapan Hidup (UHH) adalah rata-rata tahun yang diharapkan seseorang untuk hidup
HLS	Numerik	Harapan Lama Sekolah (HLS) merupakan jumlah lama pendidikan (dalam tahun) yang diharapkan dimiliki seorang anak pada usia tertentu di masa depan
RLS	Numerik	Rata-rata Lama Sekolah (RLS) menggambarkan jumlah rata-rata tahun yang dihabiskan oleh orang berusia 15 tahun ke atas di sekolah formal
PPK	Numerik	Pendapatan per Kapita Disesuaikan (PPK) merupakan jumlah pendapatan rata-rata per tahun tiap orang untuk suatu daerah
IPM	Numerik	Indeks Pembangunan Manusia (IPM) mengukur seberapa baik masyarakat dalam memanfaatkan pembangunan untuk mendapatkan hal-hal seperti uang, kesehatan, dan pendidikan.

4.2. *Pre-Processing/Data Cleaning*

Pada tahap ini penulis melakukan pengecekan *missing value* menggunakan R-Studio dan ditemukan 13 *missing value* pada data. *Missing value* tersebut kemudian ditangani menggunakan metode MICE. Setelah dipastikan tidak terdapat *missing value* pada data, langkah berikutnya adalah pengecekan *outlier*. Melalui pengecekan *outlier*, ditemukan 1 *outlier* pada atribut HLS dan 34 *outlier* pada atribut PPK. *Outlier* tersebut kemudian diatasi menggunakan metode *winsorizing* dengan memanfaatkan *Inter Quartile Range* (IQR). Dalam teknik *winsorizing* ini, setiap nilai variabel di atas atau di bawah persentil k diganti dengan nilai persentil ke-k itu sendiri. Artinya, semua nilai yang kurang dari batas bawah (persentil ke-25– 1,5 × IQR) dan lebih dari batas atas (persentil ke-75 + 1,5 × IQR) diganti dengan nilai batas bawah (persentil ke-25 – 1,5 × IQR) dan batas atas (persentil ke-75 + 1,5 × IQR) [7].

4.3. *Transformation*

Tahap tranformasi pada penelitian ini dilakukan dengan mengubah atribut kabupaten/kota yang sebelumnya bertipe data *string* menjadi data nominal sehingga atribut tersebut memiliki nilai atau *value* numerik.

4.4. *Data Mining*

4.4.1. Pemilihan Model Regresi

Pada regresi data panel, pembentukan model dilakukan dengan memilih model terbaik diantara model *Common Effect* (CEM), model *Fixed Effect* (FEM), dan model *Random Effect* (REM) melalui uji *Chow*, uji *Hausman*, dan uji *Lagrange Multiplier*. Dari hasil tiga uji tersebut, model terbaik adalah model CEM.

Tabel 2. Model data panel terbaik

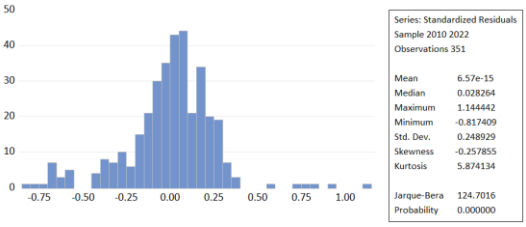
Uji Pemilihan Model	CEM	FEM	REM
Uji *Chow*	✓	×	
Uji *Hausman*		✓	×
Uji *Lagrange Multiplier*	✓		×
Model Terbaik	✓	×	×

4.4.2. Uji Asumsi Klasik

Dengan CEM sebagai model terbaik maka uji asumsi klasik perlu dilakukan agar nilai esitimasi parameter pada analisis regresi valid atau bersifat *Best Linier Unbiased Estimator* (BLUE).

1) Uji Normalitas

Dalam penelitian ini, metode *Jarque-Bera* digunakan oleh peneliti untuk menguji normalitas data.



Gambar 2. Uji normalitas

Hasil yang terlihat dalam Gambar 2 menunjukkan bahwa nilai probabilitas *Jarque-Bera* adalah 0,000000 atau kurang dari 0,05. Oleh karena itu, dapat disimpulkan bahwa data residual dalam model regresi tidak memiliki distribusi normal.

Hasil uji normalitas pada penelitian ini mengacu pada teori yang dikenal sebagai *Central Limit Theorem* (CLT), yang menyatakan bahwa data akan dianggap mengikuti distribusi normal apabila jumlah sampel yang digunakan dalam penelitian cukup besar. Menurut Ghozali dalam [8], teori CLT mensyaratkan data untuk dapat dianggap mengikuti distribusi normal ketika jumlah observasi (n) melebihi 30. Sedangkan menurut Gujarati dan Porter dalam [8], batas jumlah observasi dianggap cukup besar ketika jumlah observasi (n) lebih dari 100. Dalam penelitian ini,

terdapat 351 data observasi yang digunakan. Meskipun hasil uji normalitas menggunakan *Jarque-Berra* menunjukkan data tidak terdistribusi normal, berdasarkan asumsi CLT maka data yang dianalisis sudah memenuhi kriteria normalitas karena jumlah data observasi yang digunakan.

2) Uji Multikolinieritas

Menurut [9] untuk menentukan apakah ditemukan multikolineritas pada model regresi dapat dilakukan dengan menghitung nilai korelasi antar variabel independen, dimana jika nilai korelasi < 0,80 maka tidak ditemukan masalah multikolinearitas. Gambar 3 memperlihatkan nilai korelasi antar variabel independen penelitian ini.

Gambar 3. Uji multikolinieritas

	X1	X2	X3	X4
X1	1.000000	0.537914	0.773444	0.777032
X2	0.537914	1.000000	0.743290	0.614498
X3	0.773444	0.743290	1.000000	0.786749
X4	0.777032	0.614498	0.786749	1.000000

Dengan merujuk pada Gambar 3, didapatkan hasil yang mengindikasikan bahwa nilai korelasi dari setiap variabel independen memiliki skor kurang dari 0,80. Dengan demikian, dapat diambil kesimpulan bahwa tidak ada tanda-tanda multikolinieritas yang ditemukan di antara variabel independen dalam model regresi.

3) Uji Autokorelasi

Identifikasi autokorelasi model regresi dilakukan dengan menggunakan teknik *Durbin-Watson*. Hasil estimasi model CEM menunjukkan nilai *Durbin Watson* yang diperoleh adalah 1,957562. Berdasarkan tabel statistik, dengan menggunakan taraf signifikansi 5% diperoleh nilai Du = 1,84193 dan Dl = 1,80737. Dari nilai-nilai tersebut dapat disimpulkan bahwa $d_U\ (1{,}84193) \leq d(1{,}957562) \leq 4 - d_U\ (2{,}19263)$ yang artinya tidak ditemukan autokorelasi baik positif ataupun negatif pada model regresi.

4) Uji Heteroskedastisitas

Keputusan tentang adanya tanda-tanda heteroskedastisitas dalam model regresi dapat diambil melalui evaluasi nilai probabilitas t-statistik untuk setiap variabel. Jika nilai probabilitas t-statistik untuk tiap variabel independen melebihi ambang signifikansi (α) sebesar 5% atau 0,05, maka gejala heteroskedastisitas dalam model regresi tidak terdeteksi. Sebaliknya, jika nilai probabilitas t-statistik untuk masing-masing variabel independen berada di bawah taraf signifikansi (α) sebesar 5% atau 0,05, maka hal tersebut mengindikasikan keberadaan gejala heteroskedastisitas dalam model regresi.

Dari hasil uji yang dilakukan, terlihat bahwa nilai t-statistik untuk setiap variabel independen berada di bawah nilai signifikansi yang telah ditetapkan (0,05). Karena itu, dapat diartikan bahwa terdapat bukti adanya heteroskedastisitas dalam model regresi.

5) *Robust Standard Error*

Setelah melaksanakan uji asumsi klasik, ditemukan adanya masalah heteroskedastisitas dalam model regresi, yang berpotensi memengaruhi estimasi terhadap standar eror. Untuk menangani permasalahan heteroskedastisitas ini, salah satu solusi yang dapat diterapkan adalah penggunaan *robust standard error*.

Menurut pendapat yang diungkapkan oleh Gujarati dan Porter dalam referensi [10], *robust standard error* adalah pendekatan yang sah untuk mengatasi autokorelasi dan heteroskedastisitas dalam model regresi, tanpa menghilangkan kehadiran autokorelasi dan heteroskedastisitas tersebut. Penerapan *robust standard error* bertujuan untuk memperbaiki estimasi kesalahan standar dalam model regresi, sehingga kesimpulan statistik yang diambil tetap memiliki kevalidan yang terjaga.

Tabel 3. Perbandingan nilai *standard error* CEM OLS dengan CEM *robust*

Variabel	Standar Eror Biasa (OLS/CEM)	Standar Eror (Robust)
C	0,996327	0,0116889
UHH	0,014781	0,0230882
HLS	0,020889	0,0149042
RLS	0,018511	0,0000149
PPK	1,21E-05	0,8100619

6) Uji F

Pengambilan keputusan pada uji F dilihat berdasarkan nilai probabilitas yang diperoleh dengan taraf signifikansi 5% atau 0,05. Berdasarkan model CEM dengan *robust standard error*, nilai probabilitas yang diperoleh sebesar 0,0000 atau lebih kecil dari ambang signifikansi 0,05. Artinya, seluruh variabel independen secara bersama-sama memiliki pengaruh yang signifikan terhadap variabel dependen.

7) Uji t

Uji t dilakukan dengan membandingkan nilai probabilitas masing-masing variabel independen dengan taraf signifikansi 5% atau 0,05. Berdasarkan hasil model CEM dengan *robust standard error*, masing-masing variabel independen menunjukkan nilai probabilitas 0,0000 atau lebih kecil dari 0,05. Artinya, masing-masing variabel independen memiliki pengaruh signifikan terhadap variabel dependen.

8) Koefisien Determinasi

Adjusted R-square menunjukkan nilai sebesar 0,9978, yang mengindikasikan bahwa pengaruh variabel independen pada variabel dependen mencapai 99,78%. Artinya, hanya 0,22% yang disebabkan oleh faktor-faktor lain yang tidak diteliti. Dengan demikian, dapat diambil kesimpulan bahwa tingkat hubungan antara variabel independen dan variabel dependen bersifat sangat kuat.

9) Persamaan Regresi

Berdasarkan hasil model CEM dengan *robust standard error* diperoleh persamaan regresi sebagai berikut:

$$y' = -0{,}06262489 + 0{,}4942168_{UHH} + 1{,}9972_{HLS} + 1{,}160255_{RLS} + 0{,}0010723_{PPK}$$

10) Analisis Tren

Metode regresi data panel dapat digunakan untuk meramalkan nilai variabel dependen pada setiap individu dalam beberapa tahun mendatang, dengan ketentuan nilai variabel prediktor atau variabel independen untuk setiap individu selama tahun-tahun tersebut diketahui [11]. Oleh karena itu, untuk dapat memprediksi capaian IPM kabupaten dan kota Provinsi Jawa Barat tahun 2023, maka dilakukan proses prediksi nilai variabel independen tahun 2023 terlebih dahulu untuk kemudian hasil prediksi variabel independen tersebut disubstitusi kedalam model regresi yang telah diperoleh.

Prediksi variabel UHH, HLS, RLS, dan PPK tahun 2023 untuk masing-masing kabupaten dan kota dilakukan dengan menggunakan analisis tren linier, kuadratik, dan eksponensial. Pemilihan tren terbaik dilihat berdasarkan nilai MAPE, MAD, dan MSD yang paling kecil.

Tabel 4. MAPE, MAD, dan MSD analisis tren linier

Variabel	Tren Linear		
	MAPE	MAD	MSD
UHH	0.179847	0.128806	0.043957
HLS	1.861893	0.225347	0.079591
RLS	1.069529	0.084056	0.016039
PPK	1.769772	171.6268	60162.54
Rata-rata	1.22026	43.01626	15040.67

Tabel 5. MAPE, MAD, dan MSD analisis tren kuadratik

Variabel	Tren Kuadratik		
	MAPE	MAD	MSD
UHH	0.067389	0.048204	0.010133
HLS	2.269959	0.283489	0.764055
RLS	1.374752	0.108887	0.059003
PPK	1.483671	145.8864	40045.78
Rata-rata	1.298943	**36.58174**	**10011.65**

Tabel 6. MAPE, MAD, dan MSD analisis tren eksponensial

Variabel	Tren Eksponensial		
	MAPE	MAD	MSD
UHH	0.17824815	0.127688889	0.043752
HLS	2.03581852	0.247414815	0.106033
RLS	1.07054444	0.084407407	0.016181
PPK	2.09175556	207.1205407	135432.4
Rata-rata	1.34409167	51.89501296	33858.15

Berdasarkan hasil analisis yang diperoleh untuk masing-masing variabel independen, dapat disimpulkan bahwa analisis tren kuadratik adalah analis tren terbaik karena memiliki nilai MAD dan MSD terkecil dibandingkan analisis tren lainnya. Oleh karena itu, persamaan yang digunakan untuk memprediksi nilai variabel independen tahun 2023 adalah persamaan tren kuadratik.

11) Prediksi IPM Kabupaten/Kota Provinsi Jawa Barat Tahun 2023

Setelah mendapatkan nilai prediksi varibel independen menggunakan analisis tren kuadratik, maka dilakukan proses substitusi hasil peramalan variabel UHH, HLS, RLS, dan PPK pada tabel 4.6 kedalam model regresi yang telah diperoleh. Berikut adalah hasil prediksi capaian IPM kabupaten dan kota Provinsi Jawa Barat tahun 2023.

Tabel 7. Hasil prediksi capaian IPM

No	Kabupaten/Kota	IPM 2023
1	Bogor	85.19
2	Sukabumi	68.06
3	Cianjur	66.55
4	Bandung	74.19
5	Garut	68.35
6	Tasikmalaya	67.42
7	Ciamis	70.55
8	Kuningan	71.01
9	Cirebon	71.11
10	Majalengka	69.58
11	Sumedang	73.25
12	Indramayu	69.55
13	Subang	70.84
14	Purwakarta	72.75
15	Karawang	73.01
16	Bekasi	76.35
17	Bandung Barat	69.66
18	Pangandaran	71.97
19	Kota Bogor	78.44
20	Kota Sukabumi	76.07
21	Kota Bandung	86.20
22	Kota Cirebon	71.32
23	Kota Bekasi	85.20
24	Kota Depok	84.35
25	Kota Cimahi	79.61
26	Kota Tasikmalaya	74.90
27	Kota Banjar	73.55

4.5. *Evaluation/Knowledge*

Untuk mengetahui performa model dalam memprediksi capaian IPM kabupaten dan kota Provinsi Jawa Barat, maka dilakukan evaluasi menggunakan MAPE, MAD, dan MSD. Tabel 8 merupakan hasil evaluasi model prediksi menggunakan MAPE, MAD, dan MSD.

Tabel 8. Performa model prediksi

MAPE	MAD	MSD
0,82%	0,57	0,38

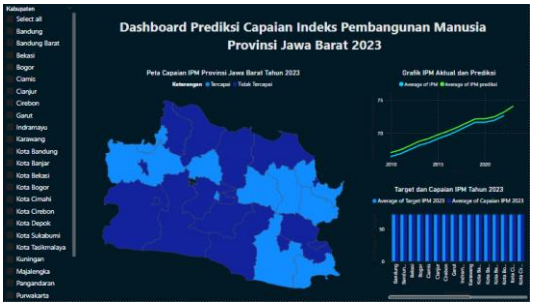
Berdasarkan Tabel 8 dapat dilihat bahwa model prediksi yang diperoleh memiliki nilai MAPE sebesar 0,82%, MAD sebesar 0,57, dan MSD sebesar 0,38. Dari hasil yang diperoleh tersebut, dapat disimpulkan bahwa model prediksi menggunakan metode regresi

data panel memiliki akurasi yang sangat baik karena nilai MAPE yang diperoleh sebesar 0,82%. Jika dilihat dari nilai MAD dan MSD, model prediksi juga bisa dikatakan layak digunakan karena nilai yang diperoleh mendekati nol.

4.6. Visualisasi Hasil

Visualisasi dilakukan agar hasil prediksi yang diperoleh dapat lebih mudah dipahami. Visualisasi dilakukan dengan membuat pemetaan capaian IPM kabupaten dan kota di Provinsi Jawa Barat tahun 2023 setelah semua tahap *modeling* selesai dan memperoleh hasil prediksi. *Tools* yang digunakan untuk membuat visualisasi pemetaan pada penelitian ini adalah *Power BI*.

Pemetaan capaian IPM kabupaten dan kota Provinsi Jawa Barat tahun 2023 dikategorikan menjadi 2, yaitu kabupaten/kota yang diprediksi dapat mencapai target IPM sesuai dengan target RPJMD (divisualisasikan dengan warna biru muda) dan kabupaten/kota yang diprediksi belum dapat mencapai target IPM sesuai dengan target RPJMD (divisualisasikan dengan warna biru tua).



Gambar 2. Pemetaan IPM kabupaten/kota Provinsi Jawa Barat

Hasil pemetaan diatas menunjukkan bahwa 11 kabupaten dan kota Provinsi Jawa Barat diprediksi mampu mencapai target IPM sesuai dengan target RPJMD, yaitu Kabupaten Bogor, Kabupaten Tasikmalaya, Kabupaten Kuningan, Kabupaten Cirebon, Kabupaten Majalengka, Kabupaten Sumedang, Kabupaten Purwakarta, Kabupaten Pangandaran, Kota Bekasi, Kota Depok, dan Kota Cimahi. Sedangkan 16 kabupaten dan kota Provinsi Jawa Barat lainnya diprediksi belum mampu mencapai target IPM sesuai dengan target RPJMD. 16 kabupaten dan kota tersebut yaitu Kabupaten Sukabumi, Kabupaten Cianjur, Kabupaten Bandung, Kabupaten Garut, Kabupaten Ciamis, Kabupaten Indramayu, Kabupaten Subang, Kabupaten Karawang, Kabupaten Bekasi, Kabupaten Bandung Barat, Kota Bogor, Kota Sukabumi, Kota Bandung, Kota Cirebon, Kota Tasikmalaya, dan Kota Banjar.

Jika dipersentasekan artinya 59,26% kabupaten dan kota Provinsi Jawa Barat masih belum dapat mencapai IPM sesuai dengan target yang ditetapkan dalam RPJMD. Oleh karena itu, agar target IPM di seluruh kabupaten/kota Provinsi Jawa Barat dapat tercapai, maka pemerintah Provinsi Jawa Barat perlu memberikan perhatian yang lebih serius pada kabupaten/kota yang masih belum mampu mencapai IPM sesuai target terutama kabupaten/kota yang memiliki selisih capaian IPM paling tinggi.

5. KESIMPULAN DAN SARAN

Dari hasil penelitian yang telah dilakukan, dapat diambil kesimpulan bahwa metode regresi data panel dapat diimplementasikan untuk memprediksi capapain IPM kabupaten/kota Provinsi Jawa Barat. Hasil prediksi yang diimplementasikan melalui pemetaan capaian IPM menggunakan Power BI menunjukkan bahwa 11 kabupaten dan kota Provinsi Jawa Barat diprediksi mampu mencapai target IPM sesuai dengan target sedangkan 16 kabupaten dan kota Provinsi Jawa Barat lainnya diprediksi belum mampu mencapai target IPM sesuai dengan target RPJMD. Akurasi model prediksi memproleh nilai MAPE sebesar 0,82%, MAD sebesar 0,57, dan MSD sebesar 0,38 sehingga model dapat dikatakan layak untuk digunakan.

Adapun untuk penelitian selanjutnya, disarankan untuk menggunakan penanganan *missing value* dan metode prediksi lain agar hasil prediksi memiliki tingkat akurasi yang lebih tinggi. Penelitian ini hanya meprediksi capaian IPM satu tahun yang akan datang sehingga untuk penelitian selanjutnya disarankan untuk melakukan prediksi capaian IPM beberapa tahun yang akan datang dan dapat menggunakan atribut lain.

DAFTAR PUSTAKA

[1] I. Kamila and U. Khairunnisa, "Perbandingan algoritma k-means dan k-medoids untuk pengelompokan data transaksi bongkar muat di provinsi riau," *Jurnal Ilmiah Rekayasa dan Manajemen Sistem Informasi*, vol. 5, no. 1, pp. 119–125, 2019.

[2] A. Fitri Boy, "Implementasi data mining dalam memprediksi harga crude palm oil (CPO) pasar domestik menggunakan algoritma regresi linier berganda (studi kasus dinas perkebunan provinsi sumatera utara)," 2020. [Online]. Available: http://jurnal.goretanpena.com/index.php/JSSR

[3] W. S *et al.*, *Data mining*. Global Eksekutif Teknologi, 2023. [Online]. Available: https://books.google.co.id/books?id=xmqvEAAAQBAJ

[4] R. Hadya, N. Begawati, and I. Yusra, "Analisis efektivitas pengendalian biaya, perputaran modal kerja, dan rentabilitas ekonomi menggunakan regresi data panel," *Jurnal Pundi*, vol. 01, no. 03, pp. 153–166, 2017.

[5] P. Tripathi, C. Bajaj, M. Bhanvadia, V. Parsekar, and V. Magar, "Comparative study of data analytics tools for effective business decision," *Vidhyayana*, vol. 8, no. 7, pp. 461–486, 2023, [Online]. Available: www.vidhyayanaejournal.org

[6] L. Addepalli, A. Lavanya, S. Sindhuja, L. Gaurav, W. Ali, and C. Author, "A comprehensive review of data visualization tools: features, strengths, and weaknesses," *IJCERT International Journal of Computer Engineering in Researvh Trends*, vol. 10, no. 1, pp. 10–20, 2023, doi: 10.22362/ijcert/2023/v10/i01/v10i0102.

[7] C. S. K. Dash, A. K. Behera, S. Dehuri, and A. Ghosh, "An outliers detection and elimination framework in classification task of data mining," *Decision Analytics Journal*, vol. 6, Mar. 2023, doi: 10.1016/j.dajour.2023.100164.

[8] D. Nur, I. Direktorat, J. Pajak, P. Fitriandi, P. Keuangan, and N. Stan, "Pengaruh kebijakan insentif pajak di masa pandemi covid-19 terhadap penerimaan PPN," 2021.

[9] G. Mardiatmoko, "Pentingnya uji asumsi klasik pada analisis regresi linier berganda," *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, vol. 14, no. 3, pp. 333–342, Oct. 2020, doi: 10.30598/barekengvol14iss3pp333-342.

[10] D. W. Utami, H. Hirawati, and A. Giovanni, "Struktur modal dan profitabilitas: studi empiris pada perusahaan sektor pertambangan periode 2014-2018," *Jurnal Ekobis: Ekonomi, Bisnis & Manajemen*, vol. 11, no. 1, pp. 186–196, 2021.

[11] Irmeilyana, I. Amalia, S. I. Maiyanti, and Ngudiantoro, "Model regresi data panel pada faktor-faktor yang menentukan produksi kopi di provinsi sumatera selatan tahun 2015-2021," *Jurnal Sains Terapan*, vol. 8, no. 1, pp. 45–56, 2022