

PERBANDINGAN METODE QUANTILE REGRESSION DAN BAYESSIAN QUANTILE REGRESSION (BQr) PADA DATA KRIMINALITAS DI JAWA TIMUR

Ade Novadio¹, Alfisyahrina Hapsery^{1*}, Artanti Indrasetianingsih¹, Fenny Fitriani¹

¹Universitas PGRI Adi Buana, Program Studi Statistika, Fakultas Teknik dan Sains
Surabaya, Indonesia.

*e-mail: alfisyahrina@unipasby.ac.id

DOI: 10.14710/j.gauss.15.1.332-343

Article Info:

Received: 2026-01-15

Accepted: 2026-08-24

Available Online: 2026-09-01

Keywords:

*Analysis Regression; Bayesian
Quantile Regression; Criminality;
Quantile Regression*

Abstract: According to 2023 Criminal Statistics, the number of crime cases in Indonesia in 2022 will reach 372,965 cases. There was a quite drastic increase in the number of crimes in 2022. The highest crime cases based on the 2023 Crime Statistics occurred in East Java, reaching 51,905 incidents in 2022. These incidents increased drastically compared to 2021, namely 19,257 incidents. There are several factors that can influence the crime rate. To determine the factors that can influence the crime rate, it is necessary to carry out an analysis using a method that can use data that contains outliers or inhomogeneous errors. The quantile regression method is a method used to estimate parameters, where this method is used on data that is not easily affected by outliers so that it does not disturb the stability of the data. Apart from using quantile regression, this research also uses the Bayesian Quantile Regression (BQr) method. By using residual values as quantile regression method analysis. The variables used in this research are crime data, percentage of population, Open Unemployment Rate, percentage of poor population, GRDP growth rate, and Labor Force Participation Rate, Human Development Index (HDI). Each data covers districts/cities in East Java in 2024.

1. PENDAHULUAN

Kriminalitas adalah segala perbuatan yang melanggar hukum pidana, norma sosial dan merugikan orang lain baik secara fisik, psikologis maupun material. Kejadian kriminalitas bukan hal baru yang terjadi di Indonesia (Chairani dkk, 2020). Menurut Laporan Statistik Kriminal 2023, jumlah kasus kejahatan di Indonesia tahun 2022 mencapai 372.965 kasus. Terjadi peningkatan jumlah kejahatan yang cukup drastis di tahun 2022, hal ini ditunjukkan dengan meningkatnya *crime rate* di tahun 2022 dari 90 di tahun 2021 menjadi 137 di tahun 2022. Sedangkan dari data Pusat Informasi Kriminal Nasional (Pusdiknas) ada sebanyak 487.753 kasus kejahatan yang terjadi di sepanjang tahun 2023. Jumlah Kriminalitas di Jawa Timur mencapai 51.905 kriminalitas pada Tahun 2022. Kejadian tersebut meningkat drastis dibandingkan tahun 2021 yaitu sebanyak 19.257 kejadian. Jika dilihat berdasarkan jumlah penduduk menurut provinsi dari data BPS tahun 2022, Jawa Timur berada di urutan kedua di Indonesia setelah Jawa Barat provinsi yang memiliki jumlah penduduk terbesar di Indonesia yaitu mencapai 41,1 juta jiwa. Jumlah kasus kriminalitas di Jawa Timur dari tahun 2021 hingga 2022 mengalami kenaikan 62,9% yaitu sebanyak 32.000 kasus. Hal ini tergolong cukup besar jika dibandingkan kasus kriminalitas yang terjadi di provinsi lainnya. Dalam beberapa kasus kejahatan dapat terjadi karena berbagai faktor. Dalam penelitian Dermawanti tahun 2015, pengangguran memiliki pengaruh signifikan secara langsung terhadap kejadian kriminalitas.

Penelitian lain menyebutkan bahwa besarnya angka pengangguran mempunyai pengaruh sosial yang luas karena tidak memiliki pekerjaan sekaligus tidak memiliki pendapatan (Kosmaryati, dkk, 2019). Semakin tinggi jumlah angka pengangguran maka

semakin tinggi pula tingkat kerawanan sosial yang ditimbulkan, salah satunya yaitu kriminalitas. Berdasarkan permasalahan tersebut, perlu dilakukan analisis untuk melihat apa saja faktor-faktor yang mempengaruhi tindak kriminalitas.

Berdasarkan data yang telah disampaikan, penelitian ini dilakukan untuk mengetahui faktor-faktor apa saja yang berpengaruh terhadap tingkat kriminalitas yang terjadi di Jawa Timur. Metode yang dapat membantu untuk memodelkan sesuai dengan sebaran data sekaligus mengetahui faktor yang mempengaruhi yaitu regresi kuantil dan bayessian regresi kuantil. Metode regresi kuantil merupakan metode yang digunakan untuk mengestimasi parameter disetiap sebaran data. Menurut Koenker, hal tersebut merupakan keuntungan utama regresi kuantil dibandingkan dengan regresi OLS.

Regresi kuantil dapat memodelkan data dalam berbagai keadaan data. Kelebihan metode ini telah dibuktikan melalui studi simulasi menggunakan Teknik *Monte Carlo*. Hasil lainnya dari simulasi ini diketahui bahwa regresi kuantil tidak hanya dapat memodelkan secara akurat pada data homogen tetapi juga memodelkan secara akurat pada data heterogen dengan variasi data beragam dan jumlah data tidak terbatas (Hapsery, dan Tribhuwaneswari, 2021). Sedangkan metode *Bayessian Quantile Regression* (BQR) yaitu gabungan antara metode regresi kuantil dan inferensi statistik Bayesian. Metode BQR ini dapat memodelkan data yang tidak seragam dengan memanfaatkan fungsi likelihood dari *asymmetric laplace Distribution* (ALD). Dalam prosesnya metode BQR menggabungkan prior untuk menghasilkan distribusi posterior. Posterior akan diperoleh dengan menggabungkan data prior (informasi parameter awal) dengan data yang akan dianalisis.

Metode BQR pernah dilakukan untuk pemodelan Indeks Pembangunan Manusia (IPM) di Jawa Tengah tahun 2022 dan memperoleh kesimpulan yaitu model yang terbentuk dari setiap kuantil pada kuantil 75% yaitu ketika jumlah penduduk bertambah 10 maka Indeks Pembangunan Manusia (IPM) di Jawa Tengah bertambah menjadi 4, sedangkan apabila tingkat kriminalitas bertambah 10 maka Indeks Pembangunan Manusia (IPM) turun sebesar 4,6 (Hapsery dkk, 2023). Berdasarkan penjelasan yang telah disampaikan, maka penelitian ini dilakukan untuk mengetahui apakah terdapat perbedaan faktor yang berpengaruh terhadap jumlah kriminalitas dengan menggunakan metode regresi kuantil dan metode *Bayessian Quantile Regression* (BQR).

2. TINJAUAN PUSTAKA

Metode yang digunakan untuk memodelkan data kriminalitas dan faktor-faktor yang mempengaruhi yaitu Regresi Linier, *Quantile Regression* dan *Bayesian Quantile Regression*. Analisis regresi membahas mengenai hubungan antara variabel yang satu dengan yang lainnya. Dengan menggeneralisirkan model regresi linier dua dan tiga variabel, model regresi variabel dependen Y dengan p variabel independen $X_1, X_2, \dots, X_p$ dapat ditulis

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_p X_{pi} + \varepsilon_i, \quad (1)$$

dengan $i = 1, 2, \dots, n$ dengan β_0 adalah *intercept*, β_1 sampai dengan β_p menyatakan *slope*, ε_i menyatakan nilai residual (Gujarati, 2004).

Regresi kuantil (*Quantile Regression*) merupakan suatu pendekatan analisis regresi yang dikenalkan oleh Koenker dan Bassett. Metode ini merupakan pengembangan dari analisis regresi yang memberikan gambaran hubungan antara variabel independen dan kuantil tertentu pada variabel dependen. Regresi kuantil adalah pendekatan analisis regresi dengan memisahkan kuantil tertentu yang memiliki nilai parameter berbeda dan dapat diterapkan pada data dengan *error* tidak homogen (Koenker, 2005).

Misal diberikan $\{y_1, y_2, y_3, \dots, y_n\}$ dan τ merupakan fungsi kumulatif atau fungsi kuantil ke- τ dari y maka

$$F_Y(y) = F(y) = P(Y \leq y) = \tau \quad (2)$$

dimana $\tau \in (0,1)$. Persamaan (2) didapatkan dari persamaan (3) yang diminimumkan fungsi terhadap $\hat{y}$ menjadi nol.

$$E(\rho(y - \hat{y})) = \int_{-\infty}^{\hat{y}} (1 - \tau)(y - \hat{y})f(y)dy + \int_{-\infty}^{\hat{y}} \tau(y - \hat{y})f(y)dy \quad (3)$$

Nilai p merupakan *loss function* dari regresi kuantil, kuantil ke- τ dari F_Y dinotasikan dengan $q_Y(\tau)$ sebagai solusi $F_Y(q) = \tau$ adalah sebagai berikut

$$q_Y(\tau) = F_Y^{-1}(\tau) = \inf\{y : F_Y(y) \geq \tau\} \quad (4)$$

dimana $\inf\{y : F_Y(y) \geq \tau\}$ merupakan fungsi invers dari $F_Y(y)$.

Persamaan umum regresi kuantil linier khusus untuk kuantil bersyarat $Q_{Y_i}(\tau | X_{1i}, X_{2i}, \dots, X_{pi})$ dari variabel dependen Y_i yaitu:

$$Y_i = \beta_0(\tau) + \beta_1(\tau)X_{1i} + \dots + \beta_p(\tau)X_{pi} + \varepsilon_i(\tau); \quad i = 1, 2, \dots, n \quad (5)$$

Apabila model regresi kuantil disajikan dalam bentuk matriks, persamaan (5) dapat ditulis sebagai persamaan (6) berikut:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & \dots & X_{p1} \\ 1 & X_{12} & \dots & X_{p2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{1n} & \dots & X_{pn} \end{bmatrix} \begin{bmatrix} \beta_0(\tau) \\ \beta_1(\tau) \\ \vdots \\ \beta_p(\tau) \end{bmatrix} + \begin{bmatrix} \varepsilon_1(\tau) \\ \varepsilon_2(\tau) \\ \vdots \\ \varepsilon_n(\tau) \end{bmatrix} \quad (6)$$

Selanjutnya persamaan (6) dapat ditulis dalam bentuk model linier berikut ini

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta}(\tau) + \boldsymbol{\varepsilon} \quad (7)$$

Jika fungsi bersyarat dari kuantil ke- τ dengan variabel X tertentu, maka fungsi bersyarat tersebut didefinisikan sebagai berikut:

$$Q_{Y_i}(\tau | X_{1i}, X_{2i}, \dots, X_{pi}) = Q_Y(\tau | X) = \mathbf{X}^T \boldsymbol{\beta}(\tau), i = 1, 2, \dots, n \quad (8)$$

Maka solusi optimasi pada regresi kuantil adalah:

$$\min_{\boldsymbol{\beta} \in \mathbf{R}^p} \sum_{i=1}^n \rho(\tau)(\mathbf{y} - \mathbf{X}^T \boldsymbol{\beta}(\tau)), i = 1, 2, \dots, n \quad \tau \in (0,1) \quad (9)$$

dimana $y_i = \{y_1, y_2, y_3, \dots, y_n\}$ merupakan sampel random dengan variabel dependen (Y) dan $x_i \in \mathbf{R}^p$ merupakan vektor kovariat, sedangkan $\rho(\tau)(\varepsilon) = \varepsilon(\tau - I(\varepsilon < 0))$ $0 < \tau < 1$ yang merupakan *loss function* yang asimetrik dari ε yang merupakan residual dari estimasi parameter (Koenker dan Hallock, 2001).

Estimasi parameter pada regresi OLS hanya dapat digunakan dalam mengatasi permasalahan *mean* atau rata-rata, oleh karena itu Koenker dan Bassett mengembangkan metode alternatif dari regresi OLS yaitu regresi kuantil. Estimasi Regresi kuantil yaitu dengan meminimalkan jumlah absolut residual yang dikenal dengan metode *Least Absolute Deviation* (LAD). Nilai *error* yang lebih besar atau sama dengan nol menggunakan bobot yaitu τ , sedangkan *error* yang kurang dari nol menggunakan pembobot $(1 - \tau)$, dan ε yang merupakan residual dari estimasi parameter (Alhamzawi dan Ali, 2020). Perkalian antara residual dengan bobot yang diberikan membentuk *loss function* ($\rho(\tau)$) yaitu,

$$\rho(\tau)(\varepsilon) = \sum_{i=1, \varepsilon_i \geq 0} \tau |\varepsilon_i| + \sum_{i=1, \varepsilon_i < 0} \tau \setminus (1 - \tau) |\varepsilon_i| \quad (10)$$

sehingga dalam regresi kuantil terdapat fungsi kuantil bersyarat ke- τ yang mempertimbangkan penduga $\hat{\beta}(\tau)$, maka diperoleh solusi untuk permasalahan tersebut yang dinyatakan sebagai berikut

$$\min_{\boldsymbol{\beta} \in \mathbf{R}^p} \sum_{i=1}^n \rho(\tau)(\mathbf{y} - Q(\tau)(y | X)); \quad \tau \in (0,1) \quad (11)$$

dengan $\rho(\tau)(\varepsilon)$ adalah *loss function* yang asimetris, τ merupakan indeks kuantil dengan $\tau \in (0,1)$ dan $Q(\tau)(y|X)$ merupakan fungsi kuantil ke- τ dari variabel Y dengan syarat X .

Parameter regresi kuantil dapat diduga dengan metode bayesian. Melalui metode *Bayesian Quantile Regression* diperlukan informasi *prior*, *likelihood*, dan distribusi *posterior*. *Loss function* dari regresi kuantil identik dengan fungsi *likelihood* dari *Asymmetric Laplace Distribution* (ALD) (Alhamzawi dan Mallick, 2020). Sehingga meminimumkan *loss function* dari regresi kuantil sama dengan memaksimalkan fungsi *likelihood* dari ALD. Diberikan variabel random ε berdistribusi ALD maka p.d.f dari $f(\varepsilon)$ yaitu,

$$f_{\tau}(\varepsilon) = \tau(1-\tau) \exp\{-\rho_{\tau}(\varepsilon)\} \quad (12)$$

dengan $0 < \tau < 1$ dan definisi $\rho_{\tau}(\varepsilon)$ yaitu $\rho_{\tau}(\varepsilon) = \varepsilon(I(\varepsilon < 0) - (1-\tau)I(\varepsilon > 0))$.

Jika $\tau = \frac{1}{2}$ maka persamaan (12) menjadi $\frac{1}{4} \exp\left(\frac{|\varepsilon|}{2}\right)$ yang merupakan bentuk standar dari distribusi *symmetric Laplace* (Yu & Moyeed, 2001). Semua nilai τ selain $\frac{1}{2}$ mengakibatkan distribusi *laplace* menjadi asimetris. Rata-rata dari ε adalah $(1-2\tau)/\tau(1-\tau)$ dan akan bernilai positif jika $\tau > \frac{1}{2}$. Variansnya adalah $(1-2\tau+2\tau^2)/\tau^2(1-\tau)^2$. Bentuk umum persamaan (12) setelah menambahkan parameter lokasi dan parameter skala (μ dan σ) yaitu:

$$f_{\tau}(\varepsilon; \mu, \sigma) = \frac{\tau(1-\tau)}{\sigma} \exp\left\{-\rho_{\tau}\left(\frac{\varepsilon-\mu}{\sigma}\right)\right\} \quad (13)$$

Estimasi bayesian untuk regresi kuantil dapat dilakukan apapun jenis distribusi dari data, dengan mengasumsikan bahwa:

1. $f(y, \mu)$ berdistribusi ALD (*Asymmetric Laplace Distribution*), dalam regresi kuantil bayesian digunakan sebagai pembentukan fungsi *likelihood* untuk pendugaan parameter.
2. $g(\mu_i) = x_i\beta(\tau) = q_{\tau}(y_i, x_i)$ untuk setiap $0 < \tau < 1$

Berdasarkan persamaan (13) jika variabel random y berdistribusi ALD maka distribusi bersama dari $y_1, y_2, \dots, y_n$

$$f(Y|X, \beta, \sigma) = \tau^n (1-\tau)^n \sigma^{-n} \exp\left\{-\sigma^{-1} \sum_{i=1}^n \rho_{\tau}(y_i - x_i\beta)\right\}. \quad (14)$$

Pemanfaatan ini memungkinkan penggunaan pendekatan normal dalam estimasi parameter dari regresi kuantil (Davino dkk, 2014). Pendekatan ini memudahkan dalam estimasi parameter dengan pendekatan *gibbs sampling*. Lemma 1 menunjukkan bahwa distribusi asimetris *laplace* setara dengan distribusi kombinasi dari distribusi normal dan eksponensial.

Lemma 1: Misalkan $u \sim \exp(1)$ dan $z \sim N(0,1)$, $\tau \in (0,1)$,

$$p = \frac{1-2\tau}{\tau(1-\tau)} \text{ dan } q = \sqrt{\frac{2}{\tau(1-\tau)}} \quad (17)$$

maka variabel random $\varepsilon \stackrel{d}{=} pu + q\sqrt{u}z$ berdistribusi ALD. Variabel dependen y_i dari lemma 1 dapat ditulis sebagai persamaan (18) berikut:

$$y_i = x_i\beta_{\tau} + pu_i + q\sqrt{u_i}z_i; \quad i = 1, 2, \dots, n. \quad (18)$$

dengan $u_i \sim \exp(1)$ dan $z_i \sim N(0,1)$. Fungsi distribusi kumulatif dari z_i didefinisikan pada persamaan (19) berikut:

$$F(z_i) = \int_{-\infty}^{u_i} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dt \quad (19)$$

Selanjutnya akan dicari distribusi y_i bersyarat u_i ,

Jika $z_i = \frac{y_i x_i \beta_\tau + p u_i}{q \sqrt{u_i}}$ dan $dz_i = \frac{1}{q \sqrt{u_i}} dy_i$ maka diperoleh:

$$F(y_i|u_i) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(y_i - x_i \beta_\tau - p u_i)^2}{2q^2 u_i}\right) \frac{1}{q \sqrt{u_i}} dy_i, \quad (20)$$

sehingga,

$$F(y_i|u_i) = \frac{1}{\sqrt{2\pi q \sqrt{u_i}}} \exp\left(-\frac{(y_i - x_i \beta_\tau - p u_i)^2}{2q^2 u_i}\right) \quad (21)$$

Menunjukkan bahwa $(y_i|u_i)$ berdistribusi normal dengan $\mu = x_i \beta_\tau - p u_i$ dan $\sigma^2 = q^2 u_i$ distribusi bersama dari sejumlah variabel random Y bersyarat U yaitu:

$$F(y_i|u_i) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi q \sqrt{u_i}}} \exp\left(-\frac{(y_i - x_i \beta_\tau - p u_i)^2}{2q^2 u_i}\right) \quad (22)$$

$$\propto \left(\prod_{i=1}^n u_i^{-\frac{1}{2}} \right) \exp\left(-\sum_{i=1}^n \frac{(y_i - x_i \beta_\tau - p u_i)^2}{2q^2 u_i}\right) \quad (23)$$

Penambahan parameter (σ) mengakibatkan persamaan (18) berubah menjadi:

$$y_i = x_i \beta_\tau + \sigma^{-1} p u_i + \sigma^{-1} q \sqrt{u_i} z_i, i = 1, 2, \dots, n. \quad (24)$$

Dalam menentukan distribusi posterior untuk parameter estimasi digunakan ALD sebagai fungsi likelihood dimana data secara langsung sulit dipecahkan secara analitis (Alhamzawi dan Ali, 2020). Untuk mengatasi kesulitan ini, digunakan pendekatan numerik dengan bantuan algoritma MCMC (*Markov Chain Monte Carlo*) yang tidak hanya efektif digunakan tetapi juga mampu mengatasi analisis yang kompleks (Lancaster, 2010) dan (Alhamzawi dan Mohammad, 2022). MSE adalah singkatan dari *Mean Squared Error*, merupakan metrik umum yang digunakan untuk mengukur rata-rata kuadrat *error*/kesalahan yang didapatkan dari rata-rata kuadrat perbedaan antara nilai prediksi dan nilai aktual (Koenker, 2005). Berikut rumus untuk MSE:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (25)$$

dimana y_i adalah nilai aktual observasi ke- i dan $\hat{y}_i$ nilai prediksi observasi ke- i yang berdasar pada hasil estimasi.

3. METODE PENELITIAN

Penelitian ini menggunakan data kriminalitas di Jawa Timur pada tahun 2024 yang diperoleh dari Badan Pusat Statistik dan data Pusiknas POLRI Kejahatan. Variabel dependen yang digunakan yaitu jumlah kriminalitas (Y). Sedangkan variabel independen yang digunakan yaitu persentase jumlah penduduk (X_1), Tingkat Pengangguran Terbuka (X_2), persentase penduduk miskin (X_3), Laju pertumbuhan PDRB (X_4), Tingkat Partisipasi Angkatan Kerja (X_5), dan Indeks Pembangunan Manusia (X_6). Tahapan analisis dalam penelitian ini yaitu:

1. Mengumpulkan dan mendeskripsikan data, memodelkan data kriminalitas dan faktor-faktor yang mempengaruhi.
2. Mengetahui karakteristik data kriminalitas dan faktor-faktor yang mempengaruhi.
3. Melakukan uji asumsi residual IIDN (Identik, Independen, dan Berdistribusi Normal) terlebih dahulu sebelum melakukan uji regresi.
4. Melakukan uji regresi kuantil yang nantinya akan diperoleh nilai MSE (*Mean Square Error*) untuk regresi kuantil.

5. Melakukan uji regresi kuantil bayessian yang nantinya akan diperoleh nilai MSE (*Mean Square Error*) untuk regresi kuantil bayessian.
6. Kemudian tahapan akhir yaitu membandingkan pendekatan mana yang menghasilkan model yang lebih baik berdasarkan kriteria nilai MSE (*Mean Square Error*) terkecil.

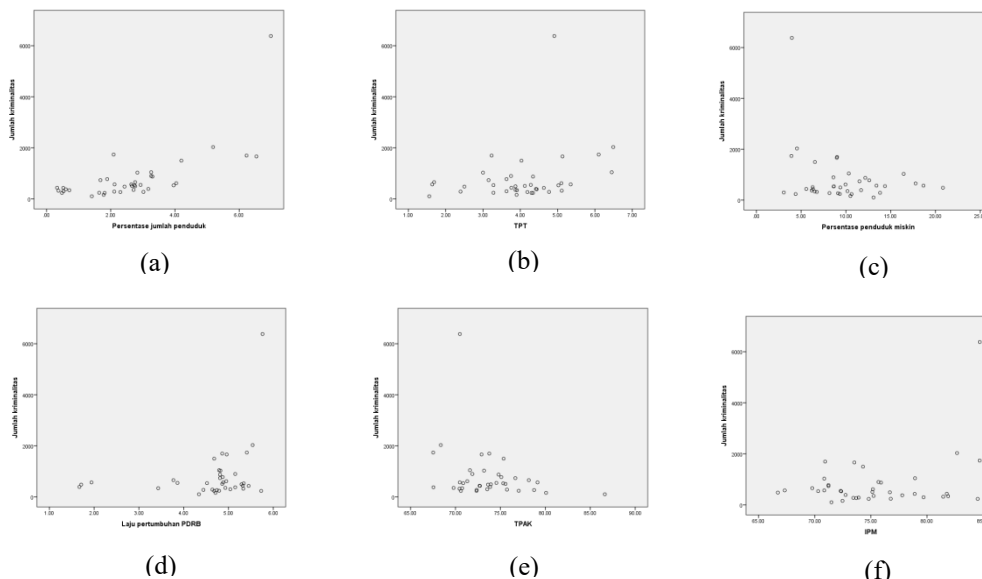
4. HASIL DAN PEMBAHASAN

Statistika deskriptif digunakan untuk mengetahui karakteristik dari variabel-variabel yang digunakan. Berikut karakteristik dari jumlah kriminalitas di Jawa Timur dan faktor-faktor yang diduga mempengaruhi. Tabel 1 menunjukkan bahwa jumlah kriminalitas di Jawa Timur tahun 2024 memiliki rata-rata sebesar 808 kasus dalam satu tahun dan daerah dengan tingkat kriminalitas tertinggi terjadi di Kota Surabaya dengan jumlah kasus sebesar 2.072 kasus pada tahun 2024. Sedangkan tingkat kriminalitas terendah terdapat pada daerah Pacitan yaitu sebesar 106 kasus dalam satu tahun. Terdapat 9 kabupaten/kota dengan tingkat kasus kriminalitas yang tinggi, serta dengan adanya nilai varians yang tinggi maka terdapat keragaman pada sebaran data kriminalitas pada kabupaten/kota di provinsi Jawa Timur.

Tabel 1. Statistika Deskriptif

Variabel	Min.	Maks.	Mean	Standar Deviasi	Varians
Y	106	6.381	808	1.046,32	1.094.788,76
X ₁	0,33	6,99	2,63	1,64	2,70
X ₂	1,56	6,49	4,04	1,18	1,39
X ₃	3,06	20,83	9,78	4,21	17,76
X ₄	1,67	5,76	4,63	0,98	0,96
X ₅	67,52	86,62	73,67	3,74	13,97
X ₆	66,72	84,69	75,31	4,76	22,64

Selanjutnya pola persebaran data antara jumlah kriminalitas beserta faktor-faktor yang diduga mempengaruhi dapat digambarkan menggunakan *scatterplot*.



Gambar 1. *Scatterplot*

Keterangan: (a) Kriminalitas dengan Persentase Jumlah Penduduk; (b) Kriminalitas dengan Tingkat Pengangguran Terbuka; (c) Kriminalitas dengan Persentase Penduduk Miskin; (d) Kriminalitas dengan Laju Pertumbuhan PDRB; (e) Kriminalitas dengan Tingkat Partisipasi Angkatan Kerja; (f) Kriminalitas dengan Indeks Pembangunan Manusia

Berdasarkan Gambar 1, diperoleh informasi bahwa pada gambar (e) menunjukkan pola data tidak simetris atau tidak menyebar secara acak. Hal ini mengindikasikan bahwa hubungan antara tingkat kriminalitas dan variabel independen tersebut sangat kecil atau tidak erat satu sama lain. Sedangkan pada gambar (a), (b), (c), (d), dan (f) data telah menyebar secara acak serta menggambarkan pola yang naik menunjukkan bahwa terdapat hubungan positif antara kriminalitas dengan masing-masing variabel independennya yang artinya apabila variabel persentase jumlah penduduk, Tingkat Pengangguran terbuka (TPT), persentase penduduk miskin, laju pertumbuhan PDRB, dan Indeks Pembangunan Manusia (IPM) tersebut mengalami kenaikan maka jumlah kriminalitas juga akan mengalami kenaikan

Analisis regresi digunakan untuk mengetahui hubungan antara variabel dependen dengan variabel independen. Hal ini perlu dilakukan sebagai penaksiran model regresi linier diperoleh dengan menggunakan estimasi OLS. Sebelum melakukan estimasi parameter regresi linier, dilakukan pemeriksaan multikolinieritas dengan menghitung nilai VIF (*Variance Inflation Factors*). Apabila terjadi multikolinieritas maka nilai $VIF > 10$. Berikut nilai VIF pada masing-masing variabel independen.

Tabel 2. Nilai *Variance Inflation Factors* (VIF)

Variabel	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆
Nilai VIF	1,27	2,27	3,84	2,02	1,97	3,93

Tabel 2 menunjukkan bahwa nilai VIF pada masing-masing variabel independen memiliki nilai kurang dari 10 yang berarti tidak terdapat multikolinearitas antar variabel independen sehingga dapat dilanjutkan dengan melakukan pengujian signifikansi parameter regresi linier. Berikut hasil estimasi parameter regresi linier yang ditampilkan pada Tabel 3.

Tabel 3. Hasil Estimasi Parameter Regresi Linier

Variabel	Koefisien	Std. Error	t-value	p-value
(Intercept)	-10.573,91	5.020,60	-2,11	0,043
X ₁	537,08	69,02	7,78	0,000
X ₂	-236,04	129,06	-1,82	0,077
X ₃	47,55	46,83	1,02	0,318
X ₄	-15,03	146,53	-0,10	0,919
X ₅	-20,26	37,81	-0,54	0,596
X ₆	159,59	41,98	3,80	0,001
Nilai F = 12,83 dengan p-value = 0,00; dan R ² = 0,713				

Tabel 3 menjelaskan mengenai uji serentak dan uji parsial. Uji serentak menghasilkan nilai *p-value* yang lebih kecil dari taraf signifikan ($\alpha = 0,05$) yang artinya terdapat minimal satu variabel yang berpengaruh secara signifikan terhadap kriminalitas di Jawa Timur. Sedangkan berdasarkan uji parsial, terdapat dua variabel yang berpengaruh signifikan terhadap kriminalitas di Jawa Timur yaitu persentase jumlah penduduk (X₁) dan Indeks Pembangunan Manusia (X₆) dimana nilai *p-value* pada masing-masing variabel independen tersebut memiliki nilai lebih kecil dari taraf signifikan ($\alpha = 0,05$). Nilai koefisien determinasi (R²) yang dihasilkan sebesar 0,713. Hal ini menunjukkan bahwa sebesar 71,3% tingkat kriminalitas di Jawa Timur mampu dijelaskan oleh model, sedangkan sisanya yaitu sebesar 28,7% dijelaskan oleh variabel lain di luar model.

Pengujian asumsi residual identik, independen, dan berdistribusi normal (IIDN) dilakukan untuk memastikan bahwa model regresi yang dihasilkan sudah valid dan akurat. Berdasarkan uji asumsi residual pada Tabel 4 menunjukkan bahwa asumsi identik belum terpenuhi karena terjadi heteroskedastisitas. Sementara itu, asumsi independen terpenuhi yang artinya tidak terdapat autokorelasi antar residual. Pada asumsi normalitas juga

terpenuhi dan disimpulkan bahwa distribusi data mengikuti distribusi normal. Hasil uji OLS, menunjukkan hanya terdapat dua variabel independen yang signifikan serta terlanggarnya asumsi Identik. Apabila dicermati dari pola data pada *scatterplot* gambar 1 terdapat pola data yang menggerombol dan menyebar yang mengindikasikan heteroskedastisitas. Hal ini diduga menjadi penyebab tidak signifikannya variabel dan terlanggarnya asumsi OLS

Tabel 4. Hasil Uji Asumsi Residual IIDN

Pengujian	Statistik uji	p-value	Keterangan
<i>Breusch Pagan Test</i>	15,808	0,007	Tidak Identik
<i>Durbin Watson Test</i>	1,613	0,000	Independen
<i>Kolmogorov Smirnov Test</i>	0,097	0,200	Berdistribusi Normal

Pada pengujian model regresi OLS terdapat pelanggaran asumsi sehingga perlu dilakukan analisis dengan model yang lebih kompleks yaitu regresi kuantil. Berikut adalah estimasi parameter dengan nilai kuantil $\tau = 0,1, 0,25, 0,5, 0,75, 0,9$ ditampilkan pada tabel 5.

Tabel 5. Hasil Estimasi Parameter β menggunakan QR pada tiap Kuantil - τ

Variabel	Parameter	Koefisien $\tau = 0,1$	Koefisien $\tau = 0,25$	Koefisien $\tau = 0,5$	Koefisien $\tau = 0,75$	Koefisien $\tau = 0,9$
Intercept	β_0	-4112,97	-2214,93	-2892,18	-2408,27	-8993,87
X_1	β_1	161,89	341,91	296,21	294,89	87,69
X_2	β_2	37,30	-146,85	-101,39	-14,55	-67,54
X_3	β_3	28,85	13,96	25,62	37,33	56,13
X_4	β_4	7,07	63,80	106,29	123,54	236,41
X_5	β_5	6,25	-18,59	-23,79	-27,79	-31,32
X_6	β_6	41,46	45,03	55,99	49,01	126,16

Tabel 5 menjelaskan bahwa parameter signifikan pada setiap kuantil berbeda, hal ini dapat diketahui dari nilai *lower CI* pada nilai parameter di setiap kuantil. Pada setiap kuantil terdapat variabel yang tidak signifikan. Pada kuantil 0,1; 0,5; dan 0,9 variabel X_2 , X_3 , X_4 , dan X_5 memiliki nilai tidak signifikan karena nilai CI melewati nol. Pada kuantil 0,25 variabel X_3 , X_4 , dan X_5 juga memiliki nilai yang tidak signifikan. Sedangkan pada kuantil 0,75 hanya variabel X_1 yang memiliki nilai signifikan karena nilai CI tidak melewati nol.

Setelah mendapatkan hasil estimasi parameter *Quantile Regression*, selanjutnya yaitu mencari nilai estimasi parameter dari *BQR*. Hasil estimasi parameter $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5$, dan β_6 menggunakan *BQR* dengan 5000 iterasi dapat dilihat pada Tabel 6 di bawah ini, dengan kuantil $\tau = 0,1, 0,25, 0,5, 0,75$, dan 0,9 serta nilai CI pada tiap parameter dan kuantil.

Tabel 6. Hasil Estimasi Parameter β menggunakan BQR pada tiap Kuantil - τ

Variabel	Parameter	Koefisien $\tau = 0,1$	Koefisien $\tau = 0,25$	Koefisien $\tau = 0,5$	Koefisien $\tau = 0,75$	Koefisien $\tau = 0,9$
Intercept	β_0	-3,08	0,08	-6,75	-4,97	-0,52
X_1	β_1	114,74	158,94	270,42	302,13	274,73
X_2	β_2	-26,08	-42,86	-66,34	-3,37	-39,40
X_3	β_3	-4,25	12,00	22,22	15,23	-4,72
X_4	β_4	-9,82	28,27	77,12	115,51	83,44
X_5	β_5	-8,48	-16,98	-37,61	-35,27	-43,80
X_6	β_6	11,45	15,84	32,59	28,73	46,44

Tabel 6 menjelaskan bahwa parameter signifikan pada setiap kuantil berbeda, hal ini dapat diketahui dari nilai *lower CI* pada nilai parameter di setiap kuantil. Pada setiap variabel di kuantil 0,1; 0,25; 0,5; 0,75; dan 0,9 memiliki nilai yang signifikan karena nilai CI tidak melewati nol. Dimana parameter signifikan pada alfa 0,05.

Model terbaik diketahui melalui nilai MSE terkecil *Quantile Regression* (QR) dan *Bayesian Quantile Regression* (BQR) seperti yang terlihat pada Tabel 7 berikut:

Quantile	MSE	
	QR	BQR
0,1	887.135,4	1.007.388
0,25	542.102,6	861.410,5
0,5	485.312,9	539.075
0,75	482.436,3	495.099,5
0,9	1.428.778	609.477,7

Tabel 7 menunjukkan bahwa model QR pada kuantil 0,1; 0,25; 0,5; dan 0,75 memiliki nilai MSE yang lebih kecil dibandingkan dengan MSE model BQR. Serta pada kuantil ke-0,75 model QR memiliki nilai MSE terkecil dibandingkan nilai MSE yang lain, sehingga dapat diartikan bahwa pendekatan *Quantile Regression* pada kuantil ke-0,75 menghasilkan model yang lebih baik dan stabil. Berikut model *Quantile Regression* (QR) pada kuantil ke-0,75.

$$\hat{y} = -2408,27 + 294,89X_1 - 14,55X_2 + 37,33X_3 + 123,54X_4 - 27,79X_5 + 49,01X_6 \quad (33)$$

Makna dari koefisien di atas adalah:

1. X_1 (persentase jumlah penduduk) memiliki nilai koefisien sebesar 294,89 menunjukkan bahwa setiap kenaikan dalam 1 poin persentase jumlah penduduk di Jawa Timur per kabupaten/kota maka jumlah kriminalitas juga akan meningkat sebesar 294,89 poin pada kuantil ke-3 0,75. Semakin tinggi persentase jumlah penduduk per kabupaten/kota, semakin meningkat pula tingkat kriminalitas yang terjadi di Jawa Timur.
2. X_2 (Tingkat Pengangguran Terbuka) memiliki nilai koefisien sebesar 14,55 menunjukkan bahwa apabila tingkat pengangguran terbuka semakin meningkat maka tingkat kriminalitas akan mengalami sedikit penurunan. Hal ini terjadi karena bukan hanya faktor tingkat pengangguran yang menjadi faktor meningkatnya kriminalitas, banyak faktor lain yang perlu dianalisis lebih lanjut. Terkadang orang yang menganggur tetapi memiliki pendidikan yang tinggi cenderung tidak terlibat dalam tindakan kriminalitas, mereka akan lebih mencari pekerjaan sampingan yang bisa menutup kebutuhan ekonomi untuk sementara.
3. X_3 (persentase penduduk miskin) memiliki nilai koefisien sebesar 37,33 hal ini menunjukkan bahwa setiap kenaikan 1 poin persentase penduduk miskin yang terjadi pada setiap kabupaten/kota di Jawa Timur maka tingkat kriminalitas juga akan meningkat sebesar 37,33 poin pada kuantil ke-3 0,75. Semakin tinggi jumlah penduduk miskin maka semakin meningkat pula tingkat kriminalitas yang terjadi di Jawa Timur. Hal ini sering terjadi karena penduduk miskin akan mengalami masalah ekonomi dan rentan mengalami tekanan sosial akibatnya akan melakukan hal apapun untuk bertahan hidup.
4. X_4 (laju pertumbuhan PDRB) memiliki nilai 123,54 yang menunjukkan bahwa setiap kenaikan 1 poin dalam laju pertumbuhan PDRB maka akan meningkatkan angka kriminalitas sebesar 123,54 poin pada kuantil ke-3 0,75 di Jawa Timur. Hal ini mungkin terjadi di wilayah yang mengalami ketimpangan ekonomi. Faktor yang lebih

berpengaruh signifikan terhadap kriminalitas bukanlah PDRB semata, melainkan ketimpangan pendapatan dan kesenjangan ekonomi. Daerah dengan PDRB tinggi namun disertai ketimpangan yang lebar (kaya dan miskin hidup berdampingan dengan kontras) sering kali mengalami tingkat kriminalitas yang lebih tinggi karena memicu rasa ketidakadilan dan tekanan ekonomi bagi kelompok berpenghasilan rendah.

5. X_5 (Tingkat Partisipasi Angkatan Kerja) memiliki nilai koefisien 27,79 yang menunjukkan bahwa apabila semakin banyak penduduk yang bekerja maka akan menurunkan nilai kriminalitas sebesar 27,79 poin di wilayah Jawa Timur. Hal ini terjadi karena semakin banyak lapangan pekerjaan yang diberikan maka akan meningkatkan pula tingkat kesejahteraan ekonomi dalam masyarakat.
6. X_6 (Indeks Pembangunan Manusia) memiliki nilai koefisien 49,01 yang menunjukkan bahwa setiap kenaikan satu poin pada IPM maka angka kriminalitas akan meningkat sebesar 49,01 poin. Dalam kondisi yang umum seharusnya apabila nilai IPM meningkat maka tingkat kriminalitas akan menurun karena IPM berpengaruh pada pendapatan, kesejahteraan sosial dan ekonomi masyarakat. Namun ada juga kondisi saat suatu wilayah mengalami ketimpangan pendapatan, ketimpangan yang tinggi bisa menimbulkan rasa ketidakadilan dan kecemburuan sosial, yang juga memicu kriminalitas.

Pemilihan model QR pada kuantil 0,75 didasarkan pada capaian nilai Mean Square Error (MSE) yang lebih rendah dibandingkan model BQR. Nilai MSE yang lebih kecil mencerminkan kemampuan model dalam menghasilkan estimasi median tingkat kriminalitas yang lebih akurat dan stabil. Keunggulan ini menunjukkan bahwa pendekatan QR lebih adaptif terhadap karakteristik data kriminalitas yang umumnya mengandung variasi distribusi, heteroskedastisitas, serta potensi outlier antar kabupaten/kota. Dengan demikian, QR terbukti menyediakan estimasi yang lebih *robust* dan representatif untuk tujuan pemodelan kriminalitas.

Hasil estimasi model menegaskan bahwa variabel persentase jumlah penduduk dan laju pertumbuhan PDRB merupakan determinan utama dalam peningkatan angka kriminalitas. Koefisien yang relatif besar pada kedua variabel tersebut menunjukkan bahwa jumlah penduduk dan angka pertumbuhan PDRB mampu mempengaruhi tingkat kriminalitas yang terjadi di setiap wilayah. Apabila laju pertumbuhan PDRB tidak merata, maka memperbesar kesenjangan ekonomi, kemiskinan, dan pengangguran yang menjadi pendorong utama kejahatan, jadi intinya, PDRB yang merata cenderung menurunkan kriminalitas, sementara PDRB yang tinggi tetapi tidak merata bisa meningkatkan kriminalitas. Serta pertumbuhan ekonomi sering kali memicu urbanisasi. Kota yang padat penduduk dengan ekonomi rendah (meski PDRB daerahnya tinggi) bisa menjadi tempat subur kejahatan karena kurangnya akses dan lapangan kerja.

Selain jumlah penduduk dan laju PDRB, variabel jumlah penduduk miskin dan IPM juga berpengaruh positif terhadap angka kriminalitas, meskipun memiliki nilai koefisien yang cenderung rendah. Hal ini mengindikasikan bahwa status kesejahteraan sosial ekonomi dalam masyarakat juga dapat mempengaruhi tingkat kriminalitas di suatu wilayah. Masyarakat yang baik secara sosial maupun ekonomi akan cenderung menurunkan jumlah tindak kejahatan pada wilayah tersebut.

Sebaliknya, variabel Tingkat Pengangguran Terbuka (TPT) dan Tingkat Partisipasi Angkatan Kerja (TPAK) menunjukkan pengaruh negatif, yang dapat diinterpretasikan sebagai indikasi bahwa tingginya angka pengangguran dan partisipasi kerja tidak selalu menunjukkan berkurangnya angka kriminalitas di suatu wilayah. Adanya faktor lain seperti faktor pendidikan, kondisi sosial, stabilitas politik, dan jaminan sosial bisa menjadi faktor

lainnya yang dapat mempengaruhi kondisi tersebut. Namun demikian akan lebih baik apabila banyaknya lapangan pekerjaan dan penyerapan jumlah pengangguran yang baik maka akan menciptakan stabilitas ekonomi dan akan menurunkan tingkat kriminalitas di suatu wilayah.

5. KESIMPULAN

Hasil penelitian menunjukkan bahwa pendekatan *Quantile Regression* atau regresi kuantil pada kuantil ke-0,75 menghasilkan model terbaik dan lebih stabil dibandingkan dengan model *Bayesian Quantile Regression* (BQR). Hal tersebut diperoleh dari nilai MSE yang terkecil dari kedua model. Model terbaik yang diperoleh yaitu terdapat pada persamaan (33). Faktor-faktor yang mempengaruhi secara signifikan terhadap tingkat kriminalitas di Jawa Timur tahun 2024 yaitu persentase jumlah penduduk (X_1), persentase penduduk miskin (X_3), laju pertumbuhan PDRB (X_4), dan IPM (X_6) di Jawa Timur.

DAFTAR PUSTAKA

- Alhamzawi, A. and G. S. Mohammad (2022). Statistical inference with normal-compound gamma priors in regression models. arXiv preprint arXiv:2212.13144.
- Alhamzawi, R. and H. T. M. Ali (2020). Brq: An r package for Bayesian quantile regression. *METRON* 78 (3), 313–328.
- Alhamzawi, R. and H. Mallick (2020). Bayesian reciprocal lasso quantile regression. *Communications in Statistics-Simulation and Computation*, 1–16.
- BPS. 2023. *Statistik Kriminal 2023*. Publikasi Badan Pusat Statistik: Jakarta, Indonesia.
- Chairani, P. A., Astuti, I.K.D., Panjaitan, N., Sinurat, S.R.Y., dkk. 2020. “Analisis Jalur pada Kejadian Kriminalitas di Indonesia Tahun 2018”, *Jurnal Sains Matematika dan Statistika*, 6(2). <https://ejournal.uin-suska.ac.id/index.php/JSMS/article/view/10555>. DOI: <https://doi.org/10.24014/jsms.v6i2.10555>.
- Davino, C., Furno, M., and Vistocco, D. *Quantile regression: Theory and Applications*, Vol. 988. John Wiley & Sons, 2014.
- Dermawanti, Hoyyi, A., Rusgiyono, A., 2015. “Faktor-Faktor yang Mempengaruhi Kriminalitas di Kabupaten Batang Tahun 2013 dengan Analisis Jalur”, *Jurnal Gaussian*, 4(2), 2015. 247-256. <https://ejournal3.undip.ac.id/index.php/gaussian/-article/view/8423>. Doi: <https://doi.org/10.14710/j.gauss.4.2.247-256>.
- Gujarati, D. N., 2004. *Basic Econometrics 4th Edition*. New York: McGraw-Hill.
- Hapsery, A., dan Tribhuwaneswari, A.B. 2021. *Monte Carlo Simulation in Quantile Regression for Sparsity Estimate. Journal of Physics: Conference Series*. <https://iopscience.iop.org/article/10.1088/1742-6596/2123/1/012027>. Doi: <https://doi.org/10.1088/1742-6596/2123/1/012027>.
- Hapsery, A., Hermanto, E. M.P., dan Aprilia, Y. U. 2023. “Bayesian Spatial Quantile Regression untuk Pemodelan Indeks Pembangunan Manusia di Jawa Tengah Tahun 2022”. *Laporan Hibah Penelitian Internal*. Universitas PGRI Adi Buana Surabaya.
- Koenker, R., dan Bassett, G. 1978, “Regression quantiles”. *Econometrica*, hal 33-50.
- Koenker, R., dan Hallock, K. F. 2001. “Quantile Regression”. *Journal of Economic Perspective*. Vol. 15, No.4, hal. 143-156. Published by: American Economic Association. <https://www.aeaweb.org/articles?id=10.1257/jep.15.4.143> Doi: <https://doi.org/10.1257/jep.15.4.143>
- Koenker, R. 2005. *Quantile Regression*. First published. Cambridge University Press.

- Kosmaryati, K., Handayani, C. A., Isfahani, R. N., Widodo, E. 2019. “Faktor-Faktor yang Mempengaruhi Kriminalitas di Indonesia Tahun 2011-2016 dengan Regresi Data Panel”, *Indonesian Journal of Applied Statistics*, 2(1). <https://jurnal-uns.ac.id/ijas/article/view/27932>. Doi: <https://doi.org/10.13057/ijas.v2i1.27932>
- Yu, K., dan Moyeed, A. A., “Bayesian quantile regression,” *Stat. Probability Letters.*, vol. 54, no. 4, pp. 437–447, 2001. <https://www.sciencedirect.com/science/article/abs/pii/S0167715201001249>
- Lancaster, T. dan Jun, S. J. 2010. “Bayesian Quantile Regression Methods”. *Journal of Applied Econometrics*, 25(4), 249–269.