



Boosted Regression Trees

Wulan Dian Pramiesti*, Fitriani Agustina, Rini Marwati

Program Studi Matematika, Fakultas Pendidikan Matematika dan Ilmu Pengetahuan Alam,
Universitas Pendidikan Indonesia, Indonesia

*Correspondence: E-mail: wdpramiesti@gmail.com

ABSTRAK

Pohon regresi merupakan teknik analisis data yang bertujuan untuk mengetahui pengaruh semua peubah penjelas terhadap peubah responnya. Namun, ternyata metode pohon regresi memiliki kelemahan yaitu struktur hierarkisnya memungkinkan terjadinya varians yang tinggi. Kelemahan metode pohon regresi ini dapat diatasi dengan menggunakan pendekatan *Boosted Regression Trees* (BRT). Pendekatan BRT menggunakan teknik *boosting* untuk menggabungkan beberapa pohon regresi secara aditif sedemikian sehingga nilai varians yang dihasilkan lebih kecil dari nilai varians sebelumnya. Penelitian ini menggunakan metode BRT untuk studi kasus menentukan faktor-faktor yang memengaruhi tingkat kriminalitas di Jawa Timur tahun 2018. Melalui 89 buah pohon optimal yang terbentuk, diperoleh kesimpulan bahwa faktor yang memengaruhi tingkat kriminalitas tertinggi adalah jumlah pemuda (31,86%), diikuti oleh persentase penduduk miskin (11,14%), Angka Partisipasi Kasar (APK) SMP (9,67%), APK SD (8,98%), kemantapan jalan (8,81%), APK SMA (8,67%), jarak ke ibukota Surabaya (7,25%), tingkat pengangguran terbuka (7,02%), Produk Domestik Regional Bruto (5,32%), dan kepadatan penduduk (1,25%).

© 2022 Kantor Jurnal dan Publikasi UPI

INFORMASI ARTIKEL

Sejarah Artikel:

Diterima 3 Juli 2021

Direvisi 10 Agustus 2021

Disetujui 6 September 2022

Tersedia Online 31 Oktober 2022

Dipublikasikan 1 Desember 2022

Kata Kunci:

Boosted Regression Trees, Boosting, Pohon Regresi, Tingkat Kriminalitas.

ABSTRACT

A regression tree is a data analysis technique that aims to determine the effect of all explanatory variables on the response variables. However, it turns out the regression tree method has a weakness that is the hierarchical structure allows for high variance predictors. The weakness of this regression tree method can be above using the Boosted Regression Trees (BRT) approach. The BRT approach uses a boosting technique to combine multiple regression trees in an additive manner so that the resulting variance is smaller than the previous variance. This study uses the BRT method for a case study of factors affecting the level of crime in East Java in 2018. Through 89 obtained optimal trees, the most important crime rate factors are the number of youth (model importance of 31.86%), followed by the percentage of poor population (11.14%), junior high school gross enrollment rate (9.67%), elementary school gross enrollment rate (8.98%), road stability (8.81%), senior high school gross enrollment rate (8.67%), distance to the capital city of Surabaya (7.25%), open unemployment rate (7.02%), Gross Domestic Regional Product (GRDP) (5.32%), and population density (1.25%).

© 2022 Kantor Jurnal dan Publikasi UPI

Keywords:

Boosted Regression Trees, Boosting, Crime Rate, Regression Trees.

1. PENDAHULUAN

Salah satu cara untuk menggali dan mendapatkan informasi lebih dari pengelolaan suatu data dapat dilakukan dengan eksplorasi data. Dalam perkembangannya analisis regresi digunakan sebagai alat untuk eksplorasi data. Analisis regresi merupakan alat statistika yang digunakan untuk mengetahui hubungan fungsional antara peubah penjelas dan peubah respon. Penaksiran nilai-nilai parameter dalam analisis regresi dilakukan dengan menggunakan Metode Kuadrat terkecil. Metode kuadrat terkecil merupakan metode penaksiran parameter analisis regresi yang paling sederhana dan umum digunakan. Namun, metode kuadrat terkecil ini memerlukan banyak asumsi yang harus dipenuhi untuk dapat dalam penaksiran parameter analisis regresi. Salah satu asumsi yang harus dipenuhinya yaitu asumsi normalitas dan asumsi linieritas (Astari et al., 2014).

Apabila kedua asumsi tersebut tidak terpenuhi, maka analisis regresi dengan penaksiran parameter menggunakan metode kuadrat terkecil tidak dapat dilakukan. Salah satu metode yang dapat dipergunakan untuk mengatasi masalah ini yaitu metode pohon regresi atau *Regression Trees*. Metode pohon regresi merupakan salah satu teknik eksplorasi data nonparametrik. Dengan kata lain, pada metode pohon regresi tidak memerlukan asumsi-asumsi tertentu. Metode pohon regresi merupakan teknik analisis data yang bertujuan untuk mengetahui pengaruh semua peubah penjelas terhadap peubah responnya dengan memperhatikan nilai peubah responnya tanpa memperhatikan asumsi-asumsi yang berlaku pada analisis regresi biasa.

Berdasarkan paparan sebelumnya, diketahui bahwa metode pohon regresi bebas asumsi, dimana hal ini merupakan salah satu keunggulan dari metode pohon regresi (Komalasari, 2007). Namun, ternyata metode pohon regresi memiliki kelemahan yaitu struktur hierarkisnya memungkinkan terjadinya *high-variance predictor* atau memiliki varians yang tinggi. Kelemahan metode pohon regresi ini dapat di atas dengan menggunakan Pendekatan *Boosted Regression Trees* (BRT). Pendekatan BRT berguna untuk mengatasi kelemahan tingginya varians dengan menggunakan teknik *boosting* untuk menggabungkan beberapa pohon regresi secara aditif untuk membentuk *low-variance predictor*. Prinsip utama dari teknik *boosting* sendiri adalah membangun dan mengkombinasikan *weak learners* dalam proses *iterative stagewise* untuk membangun *strong learners*. *Weak learners* adalah individu pada pohon regresi sedangkan *strong learners* merupakan hasil kombinasi pohon regresi dengan cara aditif.

Artikel ini membahas mengenai penggunaan BRT untuk mengetahui faktor-faktor apa saja yang memengaruhi tingkat kriminalitas di Jawa Timur. Penulis memilih lokus Jawa Timur, hal ini karena menurut liputan6.com pada tahun 2012 Provinsi Jawa Timur pernah menduduki urutan pertama dalam hal tingginya angka kejahatan. Selain itu, ternyata tahun 2019 ini tingkat kriminalitas di Jawa Timur meningkat sebesar 96%. Artikel ini akan memaparkan penerapan metode *Boosted Regression Trees* (BRT) untuk mengetahui besar pengaruh dari masing-masing faktor dalam memengaruhi tingkat kriminalitas. Metode *Boosted Regression Trees* (BRT) ini juga dapat digunakan untuk melihat bagaimana pengaruhnya secara individu terhadap tingkat kriminalitas.

2. METODOLOGI

Metode yang digunakan dan dibahas pada penelitian ini adalah metode *Boosted Regression Trees* (BRT). Data yang digunakan merupakan data sekunder yang diperoleh dari

beberapa publikasi Badan Pusat Statistik (BPS) di Provinsi Jawa Timur. Selanjutnya algoritma BRT diaplikasikan pada data yang diperoleh, lalu disajikan dalam bahasa pemrograman R.

2.1 Metode Pohon Regresi

Pohon regresi dengan p peubah penjelas $X_1, X_2, X_3, \dots, X_p$ dan satu peubah respon Y pada umumnya memiliki notasi yang sama seperti pada analisis regresi dengan p peubah penjelas $X_1, X_2, X_3, \dots, X_p$ dan satu peubah respon Y (Suharjo, 2008). Sifat dari peubah penjelas dapat berupa kategori maupun kontinu sedangkan sifat dari peubah respon untuk pohon regresi adalah kontinu.

Kriteria kehomogenan pada masing-masing simpul diketahui dengan menggunakan Jumlah Kuadrat Sisaan (JKS). Misalkan simpul t berisi anak contoh amatan dengan $n(t)$ banyaknya amatan pada simpul t dan rataan respon pada simpul t adalah (Suharjo, 2008; Lestari et al., 2022):

$$\bar{Y}(t) = \frac{1}{n(t)} \sum_{x_n \in t} Y_n$$

sehingga jumlah kuadrat sisa pada simpul t adalah:

$$JKS(t) = \sum_{x_n \in t} [Y_{i(t)} - \bar{Y}(t)]^2$$

dimana $Y_{i(t)}$ menyatakan individu peubah respon pada simpul t .

Berdasarkan Lestari et al. (2022), misalkan ada penyekatan s yang menyekat t menjadi simpul anak kiri t_L dan simpul anak kanan t_R , fungsi penyekatan untuk menentukan ukuran kehomogenan adalah:

$$\phi(s, t) = R(t) - \{R(t_L) + R(t_R)\}$$

dengan $R(t)$ adalah Jumlah Kuadrat Sisaan simpul t ($JKS(t)$), sehingga sekat terbaik s^* dari sekat yang terbentuk harus memenuhi kriteria sebagai berikut

$$\phi(s^*, t) = \max_{s \in \Omega} \phi(s, t)$$

dengan Ω adalah himpunan gugus semua kemungkinan sekat s pada simpul t .

Berdasarkan Breiman, et. al. dalam bukunya yang berjudul '*Classification and Regression Trees*' tahun 1984, penghentian pembentukan pohon dilakukan ketika banyaknya amatan pada simpul akhir kurang dari sama dengan 5. Dengan demikian proses penyekatan rekursif berakhir. Menurut Schmoor et al. (1993), menentukan banyak amatan pada simpul akhir itu adalah sebanyak kurang dari sama dengan 25 amatan.

2.2 Boosted Regression Trees

Metode *Boosted Regression Trees* (BRT) merupakan salah satu metode yang menggabungkan dua algoritma yaitu algoritma pohon regresi dan *boosting* (Elith, 2008; Zhang & Haghani, 2015, Döpke, 2017). Sebelumnya pembentukan pohon regresi dibangun dengan cara mempartisi peubah bebasnya menjadi biner secara berturut-turut berdasarkan rata-rata peubah responnya. Kemudian pohon regresi diperbaiki dengan metode *boosting* dengan membentuk pohon regresi secara rekursif mengacu pada fungsi kerugian yang dihasilkan pada pohon regresi yang sudah dibentuk sebelumnya. Metode *boosting* yang digunakan pada penelitian ini adalah metode *gradient boosting*.

Pada asumsi parameter $\{\beta_m, \mathbf{a}_m\}_1^M$ adalah titik data terdekat dengan data uji berhingga $\{y_i, \mathbf{x}_i\}_1^N$. Namun, terjadi masalah optimasi pada persamaan tersebut yakni berpontensi sulit untuk diselesaikan satu per satu dengan metode kuadrat terkecil pada persamaan $\mathbf{a}_m = \arg \min_{\mathbf{a}, \beta} \sum_{i=1}^N [-g_m(\mathbf{x}_i) - \beta h((\mathbf{x}_i; \mathbf{a}))]^2$ (Friedman, 2001). Optimasi parameter tunggal atau

ρ_m pada *gradient boosting* adalah $\rho_m = \arg \min_{\rho} \sum_{i=1}^N \psi(y_i, F_{m-1}(\mathbf{x}_i) - \rho h(\mathbf{x}_i; \mathbf{a}_m))$ yang berdasarkan fungsi kerugian atau ψ (Friedman, 2001). Pada kasus BRT $h(\mathbf{x}; \mathbf{a}_m)$ atau *base learner* berupa pohon regresi.

Pseudoresponse adalah penurunan *gradient* yang curam guna membantu optimasi parameter. *Pseudoresponse* yang digunakan pada metode BRT sama seperti pada *gradient boosting* yang dinyatakan pada persamaan (1) berikut:

$$\tilde{y}_i = -g_m(\mathbf{x}_i) = - \left[\frac{\partial \psi(y_i, F(\mathbf{x}_i))}{\partial F(\mathbf{x}_i)} \right]_{F(\mathbf{x})=F_{m-1}(\mathbf{x})} \quad (1)$$

dimana $\psi(y_i, F(\mathbf{x}_i))$ adalah fungsi kerugian dan $F_{m-1}(\mathbf{x})$ adalah nilai prediksi pada pohon $m - 1$.

Gradient boosting untuk pohon regresi menghususkan pendekatan untuk kasus dimana *base learner* $h(\mathbf{x}; \mathbf{a})$ adalah terminal simpul L pada pohon regresi. Pada setiap m iterasi, pohon regresi mempartisi ruang \mathbf{x} menjadi daerah L yang saing lepas (*disjoint*) $\{R_{lm}\}_{l=1}^L$ dan memprediksi nilai konstanta terpisah pada masing masing daerahnya yang dinyatakan pada persamaan (2) berikut.

$$h(\mathbf{x}; \{R_{lm}\}_{l=1}^L) = \sum_{i=1}^L \bar{y}_{lm} 1(\mathbf{x} \in R_{lm}) \quad (2)$$

dimana $\bar{y}_{lm} = \text{mean}_{\mathbf{x}_i \in R_{lm}}(\tilde{y}_{im})$ adalah rata-rata dari persamaan (1) di setiap wilayah R_{lm} . Parameter dari *base learner* ini adalah variabel pemisah dan titik pemisah yang sesuai untuk mendefinisikan pohon, kemudian menentukan daerah $\{R_{lm}\}_{l=1}^L$ yang berkoresponden dengan partisi pada iterasi ke- m . Dengan pohon regresi, persamaan ρ_m dapat diselesaikan secara terpisah dalam setiap region R_{lm} yang ditentukan oleh simpul terminal l yang sesuai dengan pohon ke- m . Karena pohon (2) memprediksikan nilai konstan \bar{y}_{lm} dalam setiap region R_{lm} , sehingga solusi untuk persamaan ρ_m direduksi menjadi estimasi lokal sederhana berdasarkan kriteria ψ yang bersesuaian dengan simpul terminal l dan pohon ke- m direpresentasikan sebagai persamaan (3) berikut.

$$\gamma_{lm} = \arg \min_{\gamma} \sum_{\mathbf{x}_i \in R_{lm}} \psi(y_i, F_{m-1}(\mathbf{x}_i) + \gamma) \quad (3)$$

dengan ψ menyatakan fungsi kerugian, R_{lm} menyatakan region dengan simpul terminal l dan pohon ke- m , $F_{m-1}(\mathbf{x}_i)$ menyatakan nilai prediksi pada poho $m - 1$, y_i menyatakan nilai observasi ke- i , dan γ_{lm} menyatakan *line search* pada simpul terminal l dan pohon ke- m .

Perkiraan untuk $F_{m-1}(\mathbf{x}_i)$ kemudian secara terpisah diperbarui pada masing-masing region yang berkoresponden dituliskan sebagai persamaan (4) berikut:

$$F_m = F_{m-1}(\mathbf{x}_i) + \nu \cdot \gamma_{lm} 1(\mathbf{x} \in R_{lm}) \quad (4)$$

dengan ν adalah parameter "*shrinkage*" memiliki nilai $0 < \nu \leq 1$ untuk mengontrol *learning rate* dari prosedur, secara empiris Friedman (2002) menemukan bahwa nilai kecil untuk nilai ν ($\nu \leq 1$) menyebabkan kesalahan generalisasi menjadi lebih baik. Uraian tersebut mengarah kepada algoritma untuk menggeneralisir *boosting* untuk pohon regresi adalah sebagai berikut:

$$F_0(\mathbf{x}) = \arg \min_{\rho} \sum_{i=1}^N L(y_i, \psi).$$

For $m = 1$ to M do

$$\tilde{y}_i = - \left[\frac{\partial \psi(y_i, F(\mathbf{x}_i))}{\partial F(\mathbf{x}_i)} \right]_{F(\mathbf{x})=F_{m-1}(\mathbf{x})}, i = 1, N$$

$$\{R_{lm}\}_{l=1}^L = L - \text{terminal node tree}(\{\tilde{y}_{im}, \mathbf{x}_i\}_{i=1}^N)$$

$$\gamma_{lm} = \arg \min_{\gamma} \sum_{x_i \in R_{lm}} \psi(y_i, F_{m-1}(x_i) + \gamma)$$

$$F_m(\mathbf{x}) = F_{m-1}(\mathbf{x}) + v \cdot \gamma_m \mathbf{1}(x \in R_{lm})$$

End For

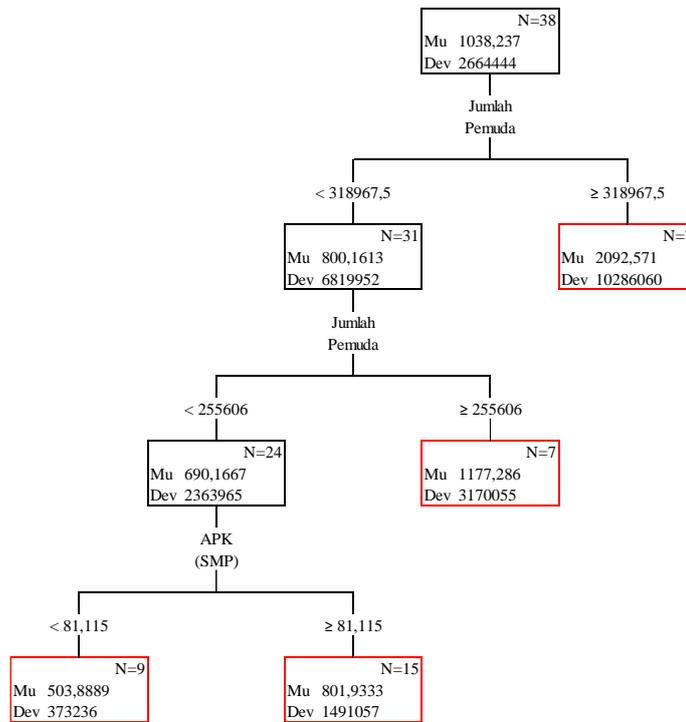
3. HASIL DAN PEMBAHASAN

Data yang digunakan pada penelitian ini adalah data sekunder yang diperoleh dari beberapa publikasi Badan Pusat Statistik (BPS) di Provinsi Jawa Timur tahun 2019. Data tersedia di *jatim.bps.go.id*. Data yang digunakan pada penelitian ini adalah data jumlah tindak pidana yang dilaporkan menurut Kabupaten/Kota sebagai peubah respon. Sebelas buah peubah penjelasnya yaitu jumlah pemuda, persentase penduduk miskin, angka partisipasi kasar SMP, angka partisipasi kasar SD, kemantapan jalan, angka partisipasi kasar SMA, jarak ke ibukota Surabaya, tingkat pengangguran terbuka, Produk Domestik Regional Bruto (PDRB), kepadatan penduduk, dan pembagian wilayah administratif. Banyaknya objek pada data sebanyak 38 objek yakni Kabupaten/Kota yang berada di Provinsi Jawa Timur.

3.1 Pohon Regresi

Pada pemisahan (*splitting*) pertama, nilai pemisah peubah penjelas jumlah pemuda di suatu wilayah terletak pada nilai 318.967,5. Hal ini mempunyai makna bahwa kabupaten/kota dengan jumlah pemuda lebih dari atau sama dengan 318.967,5 memiliki rata-rata kriminalitas lebih tinggi daripada kabupaten/kota dengan jumlah pemuda yang kurang dari 318.967,5. Sampel dengan jumlah pemudanya lebih dari atau sama dengan 318.967,5 menempati simpul akhir, dan tidak dibagi lagi menjadi simpul anakan. Simpul anakan yakni simpul yang dapat disekat menjadi dua sekatan. Hal ini karena jumlah amatannya sudah memenuhi syarat yaitu pada kasus ini jumlah amatan yang memenuhi adalah kurang dari sama dengan 15 amatan. Oleh karena itu, dilakukan penghentian pemisahan. Pada pemisahan pertama, karena jumlah amatan sebanyak 7 kurang dari 15, maka untuk sekat lebih dari atau sama dengan 318.967,5 tidak memungkinkan untuk dilakukan proses pemisahan. Pada simpul dengan jumlah pemudanya kurang dari 318.967,5 dilakukan lagi pemisahan kedua dengan peubah penjelas yang sama seperti pemisahan pertama.

Pada pemisahan kedua, nilai pemisah peubah penjelas jumlah pemuda di suatu wilayah terletak pada nilai 255.606. Kabupaten/kota dengan jumlah pemuda lebih dari atau sama dengan 255.606 menempati simpul akhir dengan jumlah amatan yang sama seperti pemisahan pertama, sehingga simpul ini tidak dapat dibagi-bagi lagi. Kemudian untuk jumlah pemuda yang kurang dari 255.606 dilakukan pemisahan ketiga dengan peubah bebas yang berbeda yakni Angka Partisipasi Kasar (APK) untuk SMP. APK pada tingkat SMP (Sekolah Menengah Pertama) yakni proporsi anak yang bersekolah SMP dengan proporsi anak yang termasuk kelompok usia yang semestinya bersekolah SMP. Peubah penjelas APK pada tingkat SMP terletak pada nilai 81,115. Nilai APK pada tingkat SMP lebih dari atau sama dengan 81,115 memiliki rata-rata kriminalitas lebih tinggi dengan rata-rata tingkat kriminalitas sebesar 801,93 sedangkan untuk nilai APK pada tingkat SMP yang kurang dari 81,115 memiliki rata-rata tingkat kriminalitas sebesar 503,89. Kedua simpulnya menempati simpul akhir karena jumlah amatannya sudah memenuhi syarat untuk dilakukan proses pemisahan lagi, sehingga pemisahan ketiga ini merupakan pemisahan terakhir karena banyak kedua simpul amatan masing-masing kurang dari atau sama dengan 15 amatan. Pohon regresi hasil perhitungan program R terdapat Gambar 1.

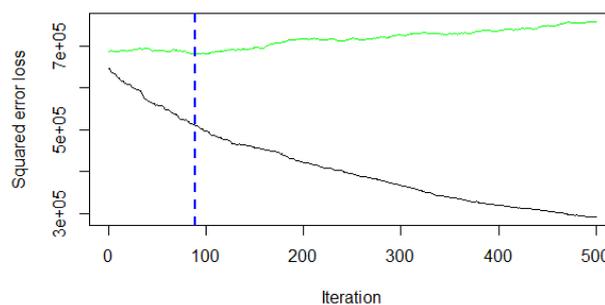


Gambar 1. Pohon Regresi Berdasarkan Hasil Keluaran Program R.

Berdasarkan pohon regresi yang terbentuk pada Gambar 1, diperoleh informasi bahwa jumlah pemuda yang menempati suatu wilayah merupakan peubah penjelas yang paling berpengaruh terhadap peubah respon tingkat kriminalitas. Perlu diingat pada penelitian ini tingkat kriminalitas diwakili oleh data jumlah tindak pidana di suatu daerah khususnya Provinsi Jawa Timur. Hal ini terbukti dari terdapatnya dua simpul terminal yang terbentuk dan keduanya dibentuk oleh peubah penjelas banyaknya jumlah pemuda di suatu wilayah. Kemudian disusul oleh peubah penjelas APK (Angka Partisipasi Kasar) SMP yang kemudian menjadi peubah kedua setelah jumlah pemuda di suatu wilayah.

3.2 Boosted Regression Trees (BRT)

Output *Boosted Regression Tree* (BRT) menghasilkan penurunan fungsi kerugian sebagai akibat dari pembentukan pohon-pohon regresi yang dibentuk dari nilai residu dari pohon sebelumnya. Grafik penurunan RMSE dan pohon optimal terdapat pada Gambar 2.



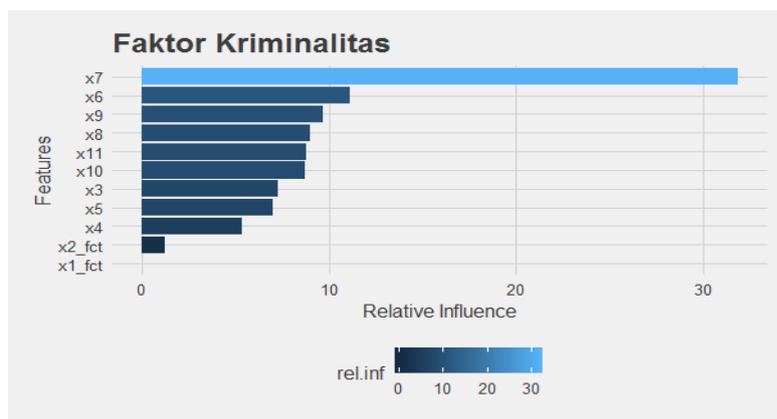
Gambar 2. Grafik Penurunan RMSE dan Pohon Optimal

Berdasarkan Gambar 2, diperoleh informasi bahwa dari pohon-pohon regresi yang terbentuk itu, pohon regresi yang terbentuk optimal berada pada iterasi ke-89. Pemilihan

pohon regresi yang terbentuk optimal ditentukan berdasarkan titik belok dan dilakukan untuk menghindari *overfitting*. Garis hitam menurun yang ada pada Gambar 2 itu merupakan fungsi kerugian yang dihasilkan dari data uji. Ketika model mulai overfit, maka grafik data tes yang berupa garis hijau pada Gambar 2 itu akan terus menaik. Dalam konteks *gradient boosting*, penghentian pohon ini didasarkan pada OOB (*out of bag*) atau CV (*cross validation*). Waktu yang ideal untuk berhenti melakukan pelatihan model adalah ketika kesalahan validasi telah menurun dan mulai stabil sebelum mulai meningkat karena *overfitting*.

Semakin banyak data yang digunakan, maka iterasi yang diperlukan juga akan semakin banyak. Hal yang menjadi faktor penentu lainnya adalah nilai *shrinkage*, dimana nilai ini merupakan representasi ketelitian dari setiap pohon yang terbentuk. Semakin kecil nilai *shrinkage*-nya, maka ketelitiannya semakin baik. Namun, kelemahannya adalah waktu yang dibutuhkan cukup lama dan iterasi yang terbentuk cukup banyak.

Pengaruh Relatif BRT juga dapat melihat berapa besar pengaruh antara masing-masing peubah penjelas terhadap peubah responnya. Besar pengaruh dari masing-masing peubah penjelas terhadap peubah responnya terdapat pada Gambar 3.



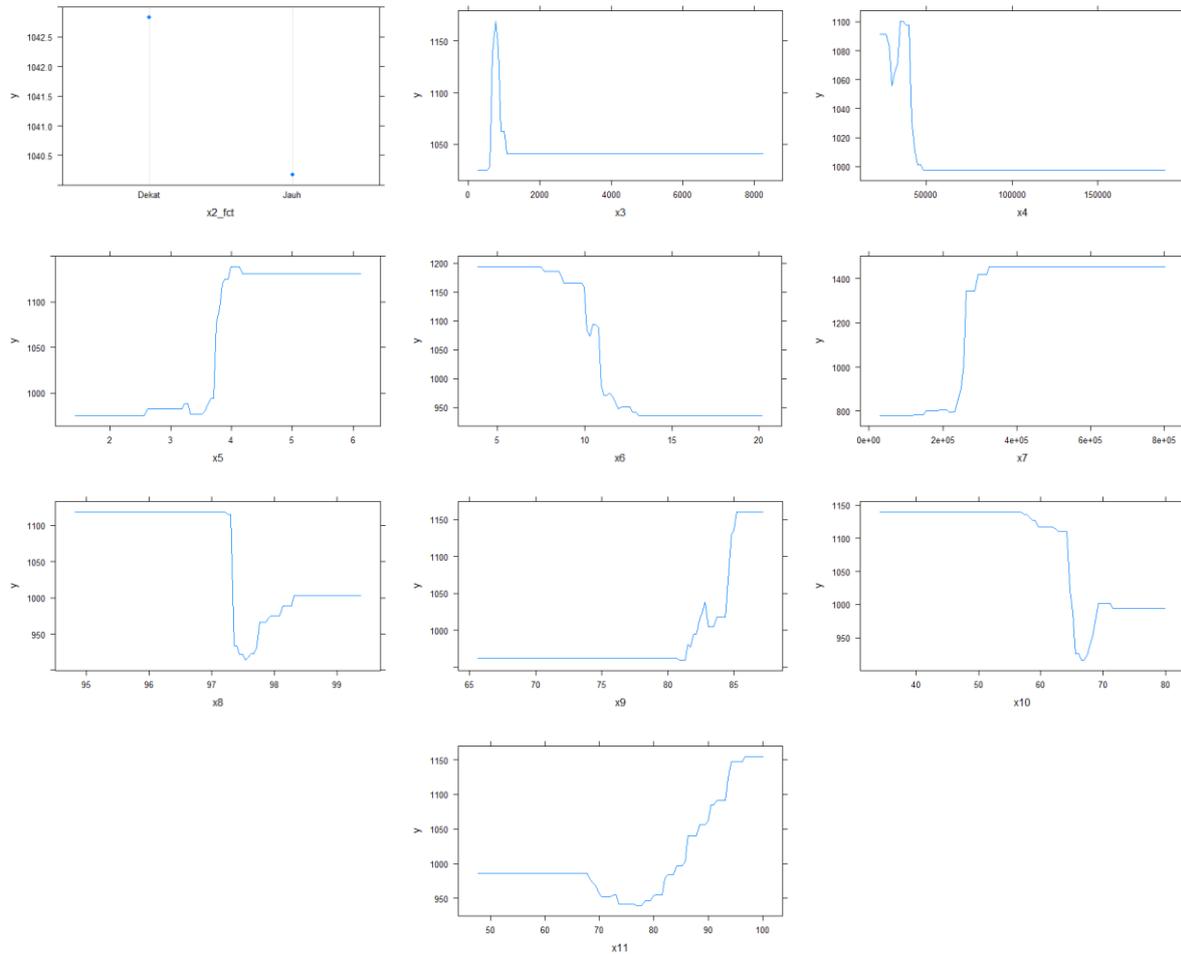
Gambar 3. Tingkat pengatuh peubah penjelas terhadap tingkat kriminalitas

Berdasarkan Gambar 3, besaran nilai pengaruh peubah penjelas terhadap peubah responnya yaitu tingkat kriminalitas dapat dideskripsikan pada Tabel 1.

Tabel 1. Persentase Peubah Penjelas yang Dihasilkan Metode BRT

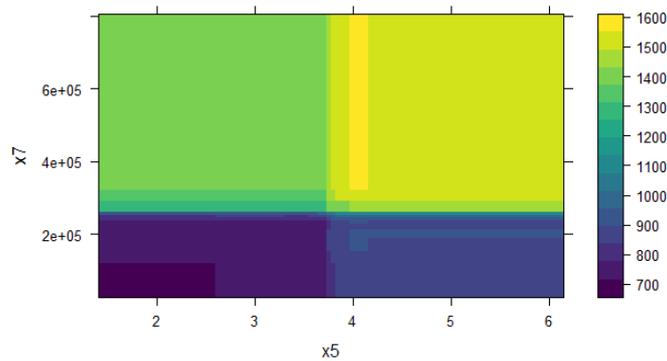
Peubah Penjelas	Persentase
Jumlah Pemuda	31.860351
Persentase Penduduk Miskin	11.14178
Angka Patisipasi Kasar SMP	9.679311
Angka Patisipasi Kasar SD	8.979713
Kemantapan Jalan	8.805536
Angka Patisipasi Kasar SMA	8.696724
Jarak ke Ibukota Surabaya	7.250867
Tingkat Pengangguran Terbuka	7.015295
PDRB	5.321855
Kepadatan Penduduk	1.248569
Pembagian Wilayah Administratif	0

Selain memperoleh informasi besar nilai masing-masing peubah penjelas terhadap peubah responnya. Metode BRT ini juga dapat melihat plot tiap peubah penjelas secara individu terhadap peubah responnya. Plot dari masing-masing peubah penjelas terhadap peubah responnya terdapat pada Gambar 4.

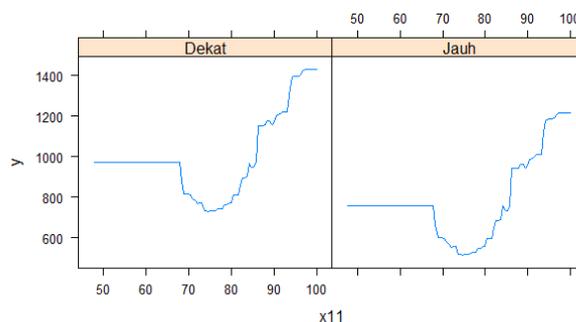


Gambar 4. Plot Pengaruh Pengaruh Masing-Masing Peubah Bebas terhadap Tingkat Kriminalitas

Metode BRT juga dapat digunakan untuk melihat pengaruh dua individu secara bersama-sama terhadap peubah responnya. Ilustrasi terdapat pada Gambar 5 dan Gambar 6.



Gambar 5. Plot Pengaruh Pengaruh Jumlah Pemuda (X_7) dan Tingkat Pengangguran Terbuka (X_5) terhadap Tingkat Kriminalitas (Y)



Gambar 6. Plot Pengaruh Jarak ke Ibukota Surabaya (X_2) dengan Kemantapan Jalan (X_{11}) terhadap Tingkat Kriminalitas (Y)

4. KESIMPULAN

Berdasarkan paparan di atas, diperoleh kesimpulan bahwa metode *Boosted Regression Tree* (BRT) merupakan suatu metode yang menggabungkan metode pohon regresi dan *boosting*, dimana metode *boosting* yang digunakan adalah *gradient boosting* kemudian konstruksi program BRT dilakukan menggunakan bahasa pemrograman R. Tahap pertama dari proses konstruksi adalah merancang program mulai dari data masukan, data keluaran dan algoritma yang kemudian diterjemahkan dalam bahasa pemrograman R. Sebelumnya dilakukan pemisahan data menjadi data uji dan data pelatihan secara acak kemudian perancangan algoritma setelah selesai maka hasil keluaran yang diperoleh kemudian diinterpretasikan.

Penerapan metode *Boosted Regression Trees* (BRT) pada studi kasus mengenai tingkat kriminalitas di Provinsi Jawa Timur, diperoleh informasi bahwa pohon regresi yang terbentuk optimal diperoleh pada iterasi ke-89. Informasi lainnya yaitu faktor yang memengaruhi tingkat kriminalitas dari yang tertinggi ke yang terendah adalah jumlah pemuda (31,86%), diikuti oleh persentase penduduk miskin (11,14%), Angka Partisipasi Kasar (APK) SMP (9,67%), APK SD (8,98%), kemantapan jalan (8,81%), APK SMA (8,67%), jarak ke ibukota Surabaya (7,25%), tingkat pengangguran terbuka (7,02%), PDRB (5,32%), dan kepadatan penduduk (1,25%).

5. DAFTAR PUSTAKA

- Astari, N. M. M., Suciptawati, N. L. P., & Sukarsa, I. K. G. (2014). Penerapan metode bootstrap residual dalam mengatasi bias pada penduga parameter analisis regresi. *E-Jurnal Matematika*, 3(4), 130-137.
- Döpke, J., Fritsche, U., & Pierdzioch, C. (2017). Predicting recessions with boosted regression trees. *International Journal of Forecasting*, 33(4), 745-759.
- Elith, J., Leathwick, J. R., & Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology*, 77(4), 802-813.
- Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29(5), 1189-1232.
- Friedman, J. H. (2002). Stochastic gradient boosting. *Computational Statistics & Data Analysis*, 38(4), 367-378.

- Komalasari, W. B. (2007). Metode pohon regresi untuk eksploratori data dengan peubah yang banyak dan kompleks. *Jurnal Informatika Pertanian*, 16(1), 967-980.
- Lestari, N. Y. T., Syafitri, U. D., & Raharjo, M. (2022). Faktor-faktor yang memengaruhi keberhasilan studi mahasiswa IPB jalur ketua OSIS dengan metode pohon regresi. *Xplore: Journal of Statistics*, 11(2), 70-83.
- Schmoor, C., Ulm, K., & Schumacher, M. (1993). Comparison of the Cox model and the regression tree procedure in analysing a randomized clinical trial. *Statistics in Medicine*, 12(24), 2351-2366.
- Suharjo, B. (2008). Metode pohon regresi dan prosedur regresi bertatar untuk segmentasi data. *Milang: Journal of Mathematics and Its Applications*, 7(2), 39-54.
- Zhang, Y., & Haghani, A. (2015). A gradient boosting method to improve travel time prediction. *Transportation Research Part C: Emerging Technologies*, 58, 308-324.