

Analisis CART (*Classification And Regression Trees*) pada Faktor-Faktor yang Mempengaruhi Kepala Rumah Tangga di Jawa Timur Melakukan Urbanisasi

Alia Hartati, Ismaini Zain dan Brodjol Sutijo Suprih Ulama

Jurusan Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Institut Teknologi Sepuluh Nopember (ITS)

Jl. Arief Rahman Hakim, Surabaya 60111

Email : brodjol_su@statistika.its.ac.id

Abstrak—Pola perkembangan kota dan urbanisasi yang pesat di kota-kota besar merupakan pertanda terjadinya kemajuan dalam pembangunan. Namun, kenyataannya urbanisasi seringkali melahirkan berbagai masalah, mulai dari masalah sosial, transportasi, hingga kriminalitas. Hasil kajian BAPPENAS menunjukkan bahwa proyeksi tingkat urbanisasi penduduk secara nasional pada tahun 2025 mencapai 68%. Proyeksi tingkat urbanisasi Jawa Timur telah mencapai angka 73,4% pada tahun 2025, hal ini menunjukkan bahwa persentase penduduk perkotaan di Jawa Timur tergolong tinggi. Penelitian ini dilakukan untuk mendeskripsikan karakteristik kepala rumah tangga Jawa Timur yang melakukan urbanisasi dan mendapatkan faktor yang mempengaruhi urbanisasi. Hasil klasifikasi dengan pendekatan CART memberikan informasi bahwa faktor-faktor yang mempengaruhi urbanisasi adalah jarak lokasi pindah, jumlah anggota rumah tangga yang ikut pindah, pendidikan tertinggi, lama waktu pindah, alasan utama pindah.

Kata Kunci—Masalah Sosial, Urbanisasi, CART, Faktor Yang Mempengaruhi Keputusan Melakukan Urbanisasi

I. PENDAHULUAN

POLA perkembangan kota yang pesat terutama di kota-kota besar merupakan pertanda terjadinya kemajuan dalam pembangunan ekonomi. Dampak dari perkembangan kota ini salah satunya adalah urbanisasi. Namun, pada kenyataannya urbanisasi seringkali melahirkan berbagai masalah, mulai dari masalah sosial, transportasi, hingga kriminalitas. Hasil kajian BAPPENAS menunjukkan bahwa berdasar proyeksi penduduk dari tahun 2005 hingga 2025 menunjukkan bahwa tingkat urbanisasi nasional diproyeksikan akan terus mengalami kenaikan dan diperkirakan akan mencapai 68% pada tahun 2025. Proyeksi tingkat urbanisasi Jawa Timur pun telah mencapai angka 73,4% pada tahun 2025, lebih tinggi dibandingkan proyeksi nasional. Hal ini menunjukkan bahwa persentase penduduk perkotaan di Jawa Timur tergolong tinggi[1].

Penelitian mengenai migrasi dan urbanisasi telah banyak dilakukan, Faktor-Faktor yang Menentukan Pilihan Daerah Tujuan Migrasi Penduduk Jawa Barat [2], Faktor-Faktor Yang Mempengaruhi Proporsi Tujuan Migrasi Keluar Dari Kalimantan Selatan [3], *Urbanization and The Poverty Level* [4], *Urbanization and The Wealth of Nations* [5].

Sementara penelitian terkait aplikasi metode CART pun telah banyak dilakukan, *Financial Application of Classification and Regression Trees* [6], Pendekatan *Bagging CART* Untuk Klasifikasi Karakteristik Anak Putus Sekolah di Jambi. Menelaah dari penelitian mengenai migrasi, urbanisasi dan aplikasi metode CART, dalam penelitian ini akan ditentukan klasifikasi faktor yang mempengaruhi kepala rumah tangga di Jawa Timur melakukan urbanisasi dengan pendekatan CART (*Classification And Regression Trees*) [7].

II. TINJAUAN PUSTAKA

A. Konsep Mobilitas Penduduk

Ada tiga komponen pertumbuhan penduduk yaitu kelahiran (fertilitas), kematian (mortalitas), dan mobilitas penduduk. Perilaku kelahiran dan kematian berbeda dengan mobilitas penduduk. Angka kelahiran dan kematian pada periode waktu tertentu mempunyai sifat stabil, tetapi mobilitas penduduk tidak ada sifat keteraturan seperti angka kelahiran dan kematian. Perilaku mobilitas penduduk adalah proses gerak penduduk dari suatu wilayah menuju wilayah lain dalam jangka waktu tertentu. Mobilitas penduduk dapat dibedakan antara mobilitas penduduk vertikal dan mobilitas penduduk horizontal. Mobilitas penduduk vertikal sering disebut dengan perubahan status, dan salah satu contohnya adalah perubahan status pekerjaan. Mobilitas penduduk horizontal, atau sering pula disebut dengan mobilitas penduduk geografis adalah gerak (*movement*) penduduk yang melintas batas wilayah menuju ke wilayah yang lain dalam periode waktu tertentu [8]. Ada pula teori yang mengatakan bahwa mengapa seseorang mengambil keputusan melakukan mobilitas, diantaranya adalah *teori kebutuhan dan stress (need and stress)* yang menyatakan bahwa proses mobilitas penduduk terjadi apabila :

- Seseorang mengalami tekanan (*stress*), baik ekonomi, sosial, maupun psikologi di tempat ia berada. Tiap-tiap individu mempunyai kebutuhan yang berbeda-beda, sehingga suatu wilayah oleh seseorang dinyatakan sebagai wilayah yang dapat memenuhi kebutuhannya, sedangkan orang lain mengatakan tidak.
- Terjadi perbedaan nilai kefaedahan wilayah antara tempat yang satu dengan tempat yang lain. Apabila tempat yang satu dengan tempat yang lain tidak ada

perbedaan nilai kefaedahan wilayah, tidak akan terjadi mobilitas penduduk.

B. Konsep Urbanisasi

Urbanisasi merupakan proses sosial penciptaan sistem dinamis yang dikenal sebagai kota. Urbanisasi meliputi perubahan penduduk, proses produksi, dan lingkungan sosial-politik-ekonomi pedesaan yang bersifat padat karya ke ekonomi kota yang terkonsentrasikan spesialisasi produksi, teknologi relatif tinggi dan kewiraswastaan. Urbanisasi dipengaruhi oleh tiga faktor yaitu pertumbuhan alami penduduk daerah perkotaan, migrasi dari daerah pedesaan ke daerah perkotaan, dan reklasifikasi desa pedesaan menjadi desa perkotaan [9]. Berhubungan erat dengan proses urbanisasi, bertambahnya kapasitas industri, intensitas pemakaian modal, teknologi, inovasi, dan spesialisasi lebih lanjut yang semuanya merupakan unsur pokok industrialisasi. Oleh karena itu, urbanisasi dipandang sebagai unsur positif berkembangnya industrialisasi di suatu tempat atau negara.

C. Klasifikasi Pohon

Classification and Regression Trees (CART) adalah salah satu metode atau algoritma dari teknik pohon keputusan. CART adalah suatu metode statistik nonparametrik yang dapat menggambarkan hubungan antara variabel respon (variabel dependen) dengan satu atau lebih variabel prediktor (variabel independen). Menurut Breiman dkk (1993), apabila variabel respon berbentuk kontinu maka metode yang digunakan adalah metode regresi pohon (*regression trees*), sedangkan apabila variabel respon memiliki skala kategorik maka metode yang digunakan adalah metode klasifikasi pohon (*classification trees*). Variabel respon dalam penelitian ini berskala kategorik, sehingga metode yang akan digunakan adalah metode klasifikasi pohon.

Pembentukan pohon klasifikasi terdiri atas 3 tahap yang memerlukan *learning sample L*. Tahap pertama adalah pemilihan pemilah. Setiap pemilihan hanya bergantung pada nilai yang berasal dari satu variabel independen [9]. Untuk variabel independen kontinu X_j dengan ruang sampel berukuran n dan terdapat n nilai amatan sampel yang berbeda, maka akan terdapat $n - 1$ pemilahan yang berbeda. Sedangkan untuk X_j adalah variabel kategori nominal bertaraf L , maka akan diperoleh pemilahan sebanyak $2^{L-1} - 1$. Tetapi jika variabel X_j adalah kategori ordinal maka akan diperoleh $L - 1$ pemilahan yang mungkin. Metode pemilahan yang sering digunakan adalah indeks Gini dengan fungsi sebagai berikut

$$i(t) = \sum_{i \neq j} p(i|t)p(j|t) \tag{1}$$

dengan $i(t)$ adalah fungsi keheterogenan indeks gini, $p(i|t)$ adalah proporsi kelas i pada simpul t , dan $p(j|t)$ adalah proporsi kelas j pada simpul t . *Goodness of split* merupakan suatu evaluasi pemilahan oleh pemilah s pada simpul t . *Goodness of split* $\phi(s, t)$ didefinisikan sebagai penurunan keheterogenan.

$$\phi(s, t) = \Delta i(s, t) = i(t) - P_L i(t_L) - P_R i(t_R) \tag{2}$$

Pengembangan pohon dilakukan dengan mencari semua kemungkinan pemilah pada simpul t_1 sehingga ditemukan pemilah s^* yang memberikan nilai penurunan keheterogenan tertinggi yaitu,

$$\Delta i(s^*, t_1) = \max_{s \in S} \Delta i(s, t_1) \tag{3}$$

dengan $\phi(s, t)$ adalah kriteria *goodness of split*, $P_L i(t_L)$ adalah proporsi pengamatan dari simpul t menuju simpul kiri, dan $P_R i(t_R)$ adalah proporsi pengamatan dari simpul t menuju simpul kanan.

Tahap kedua adalah penentuan simpul terminal. Simpul t dapat dijadikan simpul terminal jika tidak terdapat penurunan keheterogenan yang berarti pada pemilahan, hanya terdapat satu pengamatan ($n=1$) pada tiap simpul anak atau adanya batasan minimum n serta adanya batasan jumlah level atau tingkat kedalaman pohon maksimal.

Tahap ketiga adalah penandaan label tiap simpul terminal berdasar aturan jumlah anggota kelas terbanyak, yaitu:

$$p(j_0|t) = \max_j p(j|t) = \max_j \frac{N_j(t)}{N(t)} \tag{4}$$

dengan $p(j|t)$ adalah proporsi kelas j pada simpul t , $N_j(t)$ adalah jumlah pengamatan kelas j pada simpul t , dan $N(t)$ adalah jumlah pengamatan pada simpul t . Label kelas simpul terminal t adalah j_0 yang memberi nilai dugaan kesalahan pengklasifikasian simpul t terbesar.

Proses pembentukan pohon klasifikasi berhenti saat terdapat hanya satu pengamatan dalam tiap simpul anak atau adanya batasan minimum n , semua pengamatan dalam tiap simpul anak identik, dan adanya batasan jumlah level/kedalaman pohon maksimal. Setelah terbentuk pohon maksimal tahap selanjutnya adalah pemangkasan pohon untuk mencegah terbentuknya pohon klasifikasi yang berukuran sangat besar dan kompleks, sehingga diperoleh ukuran pohon yang layak berdasarkan *cost complexity pruning*, maka besarnya *resubstitution estimate* pohon T pada parameter kompleksitas α yaitu :

$$R_\alpha(T) = R(T) + \alpha |\bar{T}| \tag{5}$$

dengan $R_\alpha(T)$ adalah *resubstitution* suatu pohon T pada kompleksitas α , $R(T)$ adalah *resubstitution estimate*, α adalah *Parameter cost - complexity* bagi penambahan satu simpul akhir pada pohon T , dan $|\bar{T}|$ adalah banyaknya simpul terminal pohon T .

Cost complexity pruning menentukan pohon bagian $T(\alpha)$ yang meminimumkan $R_\alpha(T)$ pada seluruh pohon bagian untuk setiap nilai α . Nilai parameter kompleksitas α akan secara perlahan meningkat selama proses pemangkasan. Selanjutnya pencarian pohon bagian $T(\alpha) < T_{max}$ yang dapat meminimumkan $R_\alpha(T)$ yaitu :

$$R_\alpha(T(\alpha)) = \min_{T < T_{max}} R_\alpha(T) \tag{6}$$

Setelah dilakukan pemangkasan diperoleh pohon klasifikasi optimal yang berukuran sederhana namun memberikan nilai pengganti yang cukup kecil. Pengganti yang sering digunakan adalah penduga sampel uji (*test sample estimate*) dan validasi silang lipat V (*Cross Validation V-Fold Estimate*) [10].

D. Uji Dua Sampel Berpasangan

Dalam membandingkan dua perlakuan diperlukan unit eksperimen atau subjek yang memiliki kesamaan, sehingga perbedaan pada respon antara kedua grup dapat dihubungkan perbedaannya sebagai perlakuan. Dalam uji perbandingan berpasangan, respon dalam unit eksperimen dipengaruhi oleh kondisi yang terdapat dalam blok dan efek perlakuan [11]. Penelitian ini menggunakan uji hipotesis yang bersifat satu arah dengan penjelasan sebagai berikut.

$$H_0: \delta = d_0,$$

$$H_1: \delta > d_0,$$

wilayah kritis : $t > t_{(n-1, \alpha)}$

Statistik uji yang digunakan adalah sebagai berikut.

$$t = \frac{\bar{d} - D_0}{s_d / \sqrt{n}} \quad (7)$$

III. METODE PENELITIAN

Data yang digunakan dalam penelitian ini berasal dari hasil pendataan IFLS 4 (*Indonesia Family Life Survey*) pada tahun 2007. Data dihimpun oleh RAND *Labor and Population, the Center for Population and Policy Studies* (CPPS) UGM, dan Survey Meter. Total data anggota rumah tangga di Jawa Timur sebesar 1.934 data. Namun dalam penelitian hanya digunakan sebanyak 352 data dengan objek yang diteliti adalah kepala rumah tangga yang berusia lebih dari 15 tahun di Jawa Timur, pengambilan sampel penelitian ini dilakukan agar unit analisis bersifat independen antara satu dengan yang lain.

Variabel yang digunakan dalam penelitian ini adalah variabel respon dan variabel prediktor. Variabel respon berskala biner yaitu , 1 untuk kepala rumah tangga bukan pelaku urbanisasi dan 2 untuk kepala rumah tangga pelaku urbanisasi. Sedangkan variabel prediktor yang digunakan dalam penelitian ini adalah lama waktu pindah (x_1), tingkat kesejahteraan (x_2), jenjang kesejahteraan (x_2^*), jarak lokasi tempat tinggal asal ke tempat tujuan (x_3), alasan utama pindah ke tempat tujuan (x_4), jumlah anggota rumah tangga yang ikut pindah (x_5), status perkawinan (x_6), pendidikan tertinggi (x_7), dan usia saat perpindahan (x_8).

Metode analisis yang digunakan pada penelitian ini adalah analisis deskriptif untuk mengetahui karakteristik kepala rumah tangga pelaku urbanisasi di Jawa Timur. Selanjutnya, mendapatkan faktor yang mempengaruhi kepala rumah tangga di Jawa Timur berurbanisasi dengan pendekatan klasifikasi pohon (*Classification Tree*). Pembentukan klasifikasi pohon dilakukan dengan membagi data menjadi delapan kombinasi proporsi data *learning* dan *testing* yakni 95%:5%, 90%:10%, 85%:15%, 80%:20%, 75%:25%, 70%:30%, 65%:35%, dan 60%:40%. Selanjutnya dipilih satu kombinasi proporsi data *learning* dan *testing* yang memiliki ketepatan klasifikasi data *testing* terbaik untuk langkah analisis selanjutnya. Tahapan analisis klasifikasi pohon terdiri dari tiga tahapan yakni menentukan kemungkinan pemilahan pada setiap variabel prediktor, penentuan simpul terminal (penghentian pohon), dan penandaan label kelas.

IV. ANALISIS DAN PEMBAHASAN

A. Deskripsi Kepala Rumah Tangga Pelaku Urbanisasi

Analisis deskriptif memberikan informasi bahwa dari 352 sampel data, proporsi status migrasi kepala rumah tangga sebagai pelaku urbanisasi lebih sedikit (28%) dibandingkan kepala rumah tangga yang bukan urbanisasi (72%). Sebagai informasi bahwa status migrasi bukan urbanisasi tersebut terdiri dari migrasi antar desa, antar kota, dan dari kota ke desa. Sementara proporsi kepala rumah tangga berdasarkan jenis kelamin menunjukkan bahwa mayoritas kepala rumah tangga dalam kasus urbanisasi adalah laki-laki dengan persentase sebesar 81 persen.

Karakteristik terbesar kepala rumah tangga pelaku urbanisasi di Jawa Timur adalah lama waktu perpindahan kurang dari 14 tahun dengan pembagian lama waktu pindah terdiri dari jangka waktu pendek (kurang dari 14 tahun); sedang (14-28 tahun); dan panjang (28-44 tahun ke atas), tingkat kesejahteraan tergolong rendah (56%), kondisi kesejahteraan kepala rumah tangga saat urbanisasi berlangsung berada pada tingkat kesejahteraan miskin dan menengah ke bawah (52%) dan hasil uji perbandingan dua sampel menunjukkan bahwa perubahan jenjang kesejahteraan kepala rumah tangga setelah kembali ke desa dari proses urbanisasi tidak jauh berbeda dibandingkan saat urbanisasi berlangsung, jarak lokasi migrasi tidak lebih dari 350 km sehingga diduga bahwa daerah tujuan urbanisasi lebih banyak masih berada ada dalam lingkup Pulau Jawa, alasan utama pindah ke kota karena alasan pekerjaan (47%) dan keluarga (37%), jumlah anggota rumah tangga yang ikut urbanisasi memiliki jumlah yang cukup besar (1 hingga 9 orang) dan paling tidak jumlah anggota rumah tangga yang ikut pindah sebanyak 3 orang, mayoritas kepala rumah tangga telah berstatus kawin (82%), jenjang pendidikan tertinggi di bawah SMA (60%), dan usia saat perpindahan berada antara 15-64 tahun (80%).

B. Model Faktor Yang Mempengaruhi Kepala Rumah Tangga di Jawa Timur Melakukan Urbanisasi dengan Pendekatan Klasifikasi Pohon

Faktor yang mempengaruhi keputusan melakukan urbanisasi didapatkan dengan pendekatan klasifikasi pohon, di mana kategori 1 menunjukkan kategori kepala rumah tangga bukan pelaku urbanisasi dan kategori 2 menunjukkan kepala rumah tangga pelaku urbanisasi. Terlebih dahulu data penelitian dibagi menjadi dua kelompok yakni data *learning* dan data *testing*. Pada penelitian ini dicobakan kombinasi proporsi data *learning* dan *testing* antara lain 95%:5%, 90%:10%, 85%:15%, 80%:20%, 75%:25%, 70%:30%, 65%:35%, dan 60%:40%.

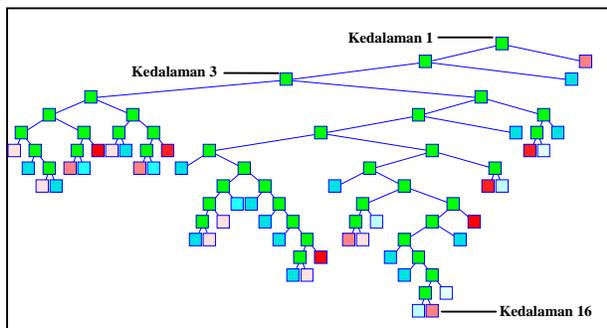
Tahap pertama pembentukan pohon klasifikasi maksimal adalah pemilihan pemilahan. Perhitungan pemilahan pada setiap variabel prediktor diperoleh hasil sebagai berikut.

- Lama waktu pindah merupakan data kontinu.
- Tingkat kesejahteraan, dengan 2 kategori ordinal. Kemungkinan pemilahan = 2-1 = 1 pemilahan.
- Jarak lokasi tempat tinggal asal ke tempat tujuan merupakan data kontinu.
- Alasan utama pindah, dengan 4 kategori nominal. Kemungkinan pemilahan = $2^{4-1} - 1 = 7$ pemilahan.
- Jumlah anggota rumah tangga yang ikut pindah merupakan data kontinu.
- Status perkawinan, dengan 2 kategori nominal. Kemungkinan pemilahan = $2^{2-1} - 1 = 1$ pemilahan.

- g. Pendidikan tertinggi, dengan 4 kategori ordinal. Kemungkinan pemilahan = $4-1 = 3$ pemilahan.
- h. Usia saat perpindahan, dengan 3 kategori nominal. Kemungkinan pemilahan = $2^{3-1} - 1 = 3$ pemilahan.

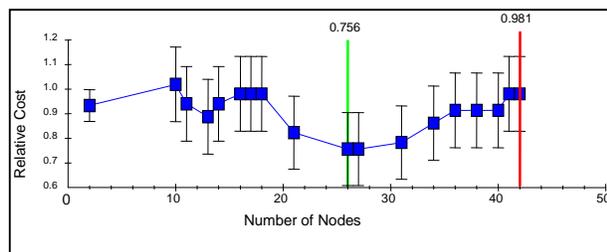
Penelitian ini menggunakan metode pemilahan Indeks Gini sesuai persamaan (1). Pemilah terbaik adalah pemilah yang menghasilkan nilai penurunan keheterogenan tertinggi (kriteria pemilahan *goodness of split* pada persamaan (2) dan persamaan (3). Pemilah terbaik pada simpul 1 (pemilah utama) pada penelitian kali ini adalah variabel jumlah anggota rumah tangga yang ikut pindah (X_5). Variabel jumlah anggota rumah tangga yang ikut pindah terpilih sebagai pemilah utama karena menghasilkan nilai penurunan keheterogenan tertinggi pada simpul 1.

Tahap kedua yaitu penentuan simpul terminal. Berdasarkan Gambar 1, pohon klasifikasi maksimal penelitian ini terdiri dari 42 simpul terminal dengan 16 kedalaman. Tahap ketiga adalah penandaan label kelas. Pemberian label kelas untuk setiap simpul terminal berdasarkan rumus pada persamaan (4). Perbedaan warna pada tiap simpul terminal menunjukkan adanya perbedaan label kelas. Ketepatan klasifikasi untuk data learning pada pohon klasifikasi maksimal ini adalah sebesar 74,25%, artinya model klasifikasi pohon maksimal yang terbentuk memiliki akurasi hasil prediksi sebesar 74,25%.



Gambar 1. Pohon Klasifikasi Maksimal

Selanjutnya dilakukan pemangkasan pohon klasifikasi maksimal. Pemangkasan pohon klasifikasi dilakukan apabila pohon klasifikasi yang terbentuk berukuran sangat besar dan kompleks dalam penggambaran struktur data sehingga pada akhirnya diperoleh ukuran pohon yang layak dan berdasarkan *cost complexity minimum* [9].

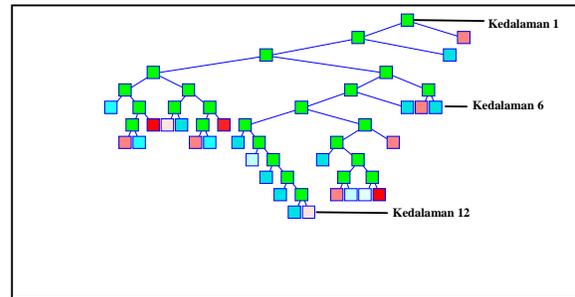


Gambar 2. Plot Relative Cost

Gambar 2 memberikan informasi bahwa nilai *relative cost* pohon klasifikasi maksimal lebih besar dibandingkan *relative cost* pohon klasifikasi optimal. Oleh karena itu perlu dilakukan pemangkasan pohon maksimal agar didapatkan nilai *relative cost* yang paling kecil. Garis hijau menunjukkan nilai *relative cost minimum* pada pohon optimal sebesar 0,756. Sedangkan untuk nilai *test set*

relative cost dan *parameter complexity* masing-masing sebesar $0,756 \pm 0,148$ dan $0,003$.

Setelah dilakukan pemangkasan terhadap pohon klasifikasi maksimal maka dihasilkan pohon klasifikasi optimal yang memiliki *relative cost* terkecil dengan 12 kedalaman dan 26 simpul terminal yang disajikan dalam Gambar 3.



Gambar 3. Pohon Klasifikasi Optimal

Variabel prediktor yang menjadi pemilah utama pada pohon klasifikasi optimal adalah jumlah anggota rumah tangga yang ikut pindah (X_5). Simpul utama (simpul 1) dipilah oleh variabel jumlah anggota rumah yang ikut pindah dari 299 pengamatan pada simpul 1, sebanyak 291 pengamatan dengan jumlah anggota rumah tangga yang ikut pindah kurang dari 4 orang dipilah ke simpul kiri (simpul 2) sedangkan sebanyak 8 jumlah anggota rumah tangga lebih dari 4 orang dipilah ke simpul kanan (simpul terminal 26) dan tidak dipilah lagi karena telah homogen dan diberi label kelas 2 atau diprediksi masuk ke dalam kategori kepala rumah tangga pelaku urbanisasi. Simpul 2 dipilah lagi berdasarkan variabel variabel pemilah jumlah anggota rumah tangga yang ikut pindah (X_5). Simpul kanan (simpul terminal 25) terdiri dari 4 pengamatan dengan jumlah anggota rumah tangga yang ikut pindah lebih dari 4 orang, simpul ini tidak dipilah lagi karena telah homogen dan diberi label kelas 1 atau diprediksi masuk ke dalam kategori kepala rumah tangga bukan pelaku urbanisasi. Sedangkan, simpul kiri (simpul 3) terdiri dari 287 pengamatan dengan jumlah anggota rumah tangga yang ikut pindah kurang dari 4 orang. Pemilahan akan dilakukan terus-menerus hingga terbentuk simpul terminal.

Dari hasil pemilahan tersebut, didapatkan 26 simpul terminal. Karakteristik kepala rumah tangga pada kasus urbanisasi terbentuk pada simpul terminal 2, 4, 5, 7, 9, 10, 17, 20, 21, 23, dan 26. Dari kesebelas simpul terminal yang terbentuk tersebut, maka dapat diketahui karakteristik kepala rumah tangga pelaku urbanisasi. Berikut ini merupakan penjelasan rinci dari simpul-simpul terminal yang terbentuk.

- **Simpul terminal 1** (49 pengamatan) diprediksi sebagai kelompok kepala rumah tangga bukan pelaku urbanisasi. Karakteristik kepala rumah tangga dari simpul ini adalah alasan utama pindah ke tempat tujuan sehubungan dengan keluarga dan lainnya, jarak lokasi pindah kurang dari 390 km, dan jumlah anggota rumah tangga yang ikut pindah kurang dari 1 orang.
- **Simpul terminal 2** (4 pengamatan) diprediksi sebagai kepala rumah tangga pelaku urbanisasi. Karakteristik kepala rumah tangga dari simpul ini adalah jarak lokasi pindah kurang dari 105 km, jenjang pendidikan tertinggi adalah tidak sekolah dan di bawah SMA, alasan utama pindah sehubungan dengan pekerjaan non militer dan

- pendidikan/latihan, jumlah anggota rumah tangga yang ikut pindah antara 2 hingga 4 orang.
- **Simpul terminal 3** (12 pengamatan) diprediksi sebagai kepala rumah tangga bukan pelaku urbanisasi. Karakteristik kepala rumah tangga dari simpul ini adalah jumlah anggota rumah tangga antara 2 hingga 4 orang, alasan utama pindah sehubungan dengan pekerjaan non militer dan pendidikan/latihan, jenjang pendidikan tertinggi adalah tidak sekolah dan di bawah SMA, dan jarak lokasi pindah antara 105 km hingga 390 km.
 - **Simpul terminal 4** (7 pengamatan) diprediksi sebagai kepala rumah tangga pelaku urbanisasi. Karakteristik kepala rumah tangga dari simpul ini adalah alasan utama pindah sehubungan dengan pekerjaan (non militer) dan pendidikan/latihan, jumlah anggota rumah tangga yang ikut pindah antara 2 hingga 4 orang, dan jenjang pendidikan tertinggi merupakan kategori SMA dan di atas SMA, dan jarak lokasi pindah kurang dari 390 km.
 - **Simpul terminal 5** (9 pengamatan) diprediksi sebagai kepala rumah tangga pelaku urbanisasi. Karakteristik kepala rumah tangga dari simpul ini adalah jumlah anggota rumah tangga yang ikut pindah antara 2 hingga 4 orang, jarak lokasi pindah antara 390 km hingga 550 km dan jenjang pendidikan tertinggi merupakan kategori tidak sekolah, SMA dan di atas SMA.
 - **Simpul terminal 6** (11 pengamatan) diprediksi sebagai kepala rumah tangga bukan pelaku urbanisasi. Karakteristik kepala rumah tangga dari simpul ini adalah jarak lokasi pindah lebih dari 550 km, jenjang pendidikan terakhir adalah di bawah SMA, dan jumlah anggota rumah tangga yang ikut pindah kurang dari 2 orang.
 - **Simpul terminal 7** (4 pengamatan) diprediksi sebagai kepala rumah tangga pelaku urbanisasi. Karakteristik kepala rumah tangga dari simpul ini adalah tingkat kesejahteraan tergolong dalam kategori tinggi, jarak lokasi pindah antara 390 km hingga 800 km, jumlah anggota rumah tangga yang ikut pindah kurang dari 2 orang.
 - **Simpul terminal 8** (8 pengamatan) diprediksi sebagai kepala rumah tangga bukan pelaku urbanisasi. Karakteristik kepala rumah tangga dari simpul ini adalah tingkat kesejahteraan termasuk kategori tinggi, jarak pindah lebih dari 800 km, jumlah anggota rumah tangga yang ikut pindah kurang dari 2 orang, dan jenjang pendidikan tertinggi termasuk kategori di bawah SMA.
 - **Simpul terminal 9** (11 pengamatan) diprediksi sebagai kepala rumah tangga pelaku urbanisasi. Karakteristik kepala rumah tangga dari simpul ini adalah tingkat kesejahteraan rendah, jarak lokasi pindah lebih dari 390 km, jumlah anggota rumah tangga yang ikut pindah antara 2 hingga 4 orang, dan jenjang pendidikan tertinggi adalah di bawah SMA.
 - **Simpul terminal 10** (2 pengamatan) diprediksi sebagai kepala rumah tangga bukan pelaku urbanisasi. Karakteristik kepala rumah tangga dari simpul ini adalah lama waktu tinggal kurang dari 3 tahun, jarak lokasi pindah kurang dari 38 km, dan jumlah anggota rumah tangga yang ikut pindah antara 2 hingga 4 orang.
 - **Simpul terminal 11** (35 pengamatan) diprediksi sebagai kepala rumah tangga bukan pelaku urbanisasi. Karakteristik kepala rumah tangga dari simpul ini adalah lama waktu tinggal kurang dari 3 tahun, jarak lokasi pindah kurang dari 38 km, dan jumlah anggota rumah tangga yang ikut pindah antara 2 hingga 4 orang.
 - **Simpul terminal 12** (10 pengamatan) diprediksi sebagai kepala rumah tangga bukan pelaku urbanisasi. Karakteristik kepala rumah tangga dari simpul ini adalah lama waktu tinggal kurang dari 3 tahun, jarak lokasi pindah kurang dari 38 km, dan jumlah anggota rumah tangga yang ikut pindah antara 2 hingga 4 orang.
 - **Simpul terminal 13** (3 pengamatan) diprediksi sebagai kepala rumah tangga bukan pelaku urbanisasi. Karakteristik kepala rumah tangga dari simpul ini adalah lama waktu tinggal kurang dari 3 tahun, jarak lokasi pindah kurang dari 38 km, jumlah anggota rumah tangga yang ikut pindah antara 2 hingga 4 orang, dan usia saat pindah merupakan kategori 15-64 tahun.
 - **Simpul terminal 14** (2 pengamatan) diprediksi sebagai kepala rumah tangga bukan pelaku urbanisasi. Karakteristik kepala rumah tangga dari simpul ini adalah lama waktu tinggal kurang dari 3 tahun, jarak lokasi pindah kurang dari 38 km, alasan utama pindah sehubungan dengan keluarga dan lainnya, jumlah anggota rumah tangga yang ikut pindah antara 2 hingga 4 orang, dan usia saat pindah merupakan kategori 15-64 tahun.
 - **Simpul terminal 15** (14 pengamatan) diprediksi sebagai kepala rumah tangga pelaku urbanisasi. Karakteristik kepala rumah tangga dari simpul ini adalah lama waktu tinggal kurang dari 3 tahun, jarak lokasi pindah kurang dari 38 km, alasan utama pindah sehubungan dengan pekerjaan non militer dan pendidikan/latihan, jumlah anggota rumah tangga yang ikut pindah antara 2 hingga 4 orang, dan usia saat pindah termasuk kategori 15-64 tahun.
 - **Simpul terminal 16** (8 pengamatan) diprediksi sebagai kepala rumah tangga bukan pelaku urbanisasi. Karakteristik kepala rumah tangga dari simpul ini adalah lama waktu tinggal lebih dari 3 tahun, jarak lokasi pindah kurang dari 22 km, jumlah anggota rumah tangga yang ikut pindah antara 2 hingga 4 orang, dan jenjang pendidikan tertinggi adalah tidak sekolah, di bawah SMA, dan di atas SMA.
 - **Simpul terminal 17** (17 pengamatan) diprediksi sebagai kepala rumah tangga pelaku urbanisasi. Karakteristik kepala rumah tangga dari simpul ini adalah lama waktu tinggal kurang dari 7 tahun, jarak lokasi pindah kurang dari 38 km, jumlah anggota rumah tangga yang ikut pindah antara 2 hingga 4 orang, dan jenjang pendidikan tertinggi merupakan kategori tidak sekolah, di bawah SMA, dan di atas SMA.
 - **Simpul terminal 18** (6 pengamatan) diprediksi sebagai kepala rumah tangga bukan pelaku urbanisasi. Karakteristik kepala rumah tangga dari simpul ini adalah lama waktu tinggal lebih dari 7 tahun, jarak lokasi pindah kurang dari 38 km, jumlah anggota rumah tangga yang ikut pindah antara 2 hingga 4 orang, dan jenjang pendidikan tertinggi merupakan kategori tidak sekolah, di bawah SMA, dan di atas SMA.
 - **Simpul terminal 19** (21 pengamatan) diprediksi sebagai kepala rumah tangga bukan pelaku urbanisasi. Karakteristik kepala rumah tangga dari simpul ini adalah jarak lokasi pindah kurang dari 38 km, jumlah anggota rumah tangga yang ikut pindah antara 2 hingga 4 orang,

dan jenjang pendidikan tertinggi merupakan kategori tidak sekolah, di bawah SMA, dan di atas SMA.

- **Simpul terminal 20** (1 pengamatan) diprediksi sebagai kepala rumah tangga pelaku urbanisasi. Karakteristik kepala rumah tangga dari simpul ini adalah lama waktu tinggal lebih dari 3 tahun, jarak lokasi pindah kurang dari 38 km, dan jumlah anggota rumah tangga yang ikut pindah antara 2 hingga 4 orang, dan jenjang pendidikan tertinggi merupakan kategori tidak sekolah, di bawah SMA, dan di atas SMA.
- **Simpul terminal 21** (25 pengamatan) diprediksi sebagai kepala rumah tangga pelaku urbanisasi. Karakteristik kepala rumah tangga dari simpul ini adalah lama waktu tinggal lebih dari 3 tahun, jarak lokasi pindah kurang dari 38 km, jumlah anggota rumah tangga yang ikut pindah antara 2 hingga 4 orang, dan jenjang pendidikan merupakan kategori SMA.
- **Simpul terminal 22** (13 pengamatan) diprediksi sebagai kepala rumah tangga bukan pelaku urbanisasi. Karakteristik kepala rumah tangga dari simpul ini adalah jarak lokasi pindah kurang dari 38 km dan jumlah anggota rumah tangga yang ikut pindah antara 2 hingga 4 orang.
- **Simpul terminal 23** (13 pengamatan) diprediksi sebagai kepala rumah tangga pelaku urbanisasi. Karakteristik kepala rumah tangga dari simpul ini adalah jarak lokasi pindah antara 38 km hingga 475 km dan jumlah anggota rumah tangga yang ikut pindah antara 2 hingga 4 orang.
- **Simpul terminal 24** (2 pengamatan) diprediksi sebagai kepala rumah tangga bukan pelaku urbanisasi. Karakteristik kepala rumah tangga dari simpul ini adalah jarak lokasi pindah lebih dari 475 km dan jumlah anggota rumah tangga yang ikut pindah antara 2 hingga 4 orang.
- **Simpul terminal 25** (4 pengamatan) diprediksi sebagai kepala rumah tangga bukan pelaku urbanisasi. Karakteristik kepala rumah tangga dari simpul ini adalah jumlah anggota rumah tangga yang ikut pindah antara 2 hingga 4 orang.
- **Simpul terminal 26** (8 pengamatan) diprediksi sebagai kepala rumah tangga pelaku urbanisasi. Karakteristik kepala rumah tangga dari simpul ini adalah jumlah anggota rumah tangga yang ikut pindah lebih dari 4 orang.

Ketepatan klasifikasi data *learning* dari pohon optimal pada penelitian ini sebesar 75,92%. Validasi pohon optimal dilakukan dengan memasukkan data *testing* ke dalam model pohon klasifikasi yang terbentuk dari data *learning* hingga akhirnya diperoleh ketepatan klasifikasi sebesar 66,04%. Dari kombinasi antara data *learning* dan data *testing* pada Tabel 1, ketepatan klasifikasi data *testing* tertinggi dicapai pada kombinasi data *learning* 85% dan *testing* 15% dengan jumlah terminal sebanyak 26 simpul, sehingga kombinasi inilah yang digunakan sebagai acuan analisis.

Tabel 1. Perbandingan Ketepatan Klasifikasi Antar Kombinasi Data

No	Kombinasi Data (%)		Ketepatan Klasifikasi (%)		Jumlah Node Terminal
	<i>Learning</i>	<i>Testing</i>	<i>Learning</i>	<i>Testing</i>	
1	95	5	77,84	55,55	54
2	90	10	75,70	57,14	45
3*	85	15	75,92	66,04	26
4	80	20	75,45	52,11	32
5	75	25	70,12	55,17	20
6	70	30	73,28	55,24	25
7	65	35	76,41	52,03	30
8	60	40	67,77	52,48	9

Keterangan : * : kombinasi data yang digunakan analisis selanjutnya

V. KESIMPULAN

Berdasarkan analisis dan pembahasan, didapatkan kesimpulan bahwa karakteristik kepala rumah tangga pada kasus urbanisasi di Jawa Timur mayoritas memiliki lama waktu pindah kurang dari 14 tahun, tingkat kesejahteraan kepala rumah tangga termasuk dalam kategori rendah dan kondisi kesejahteraan setelah urbanisasi tidak jauh berbeda dibandingkan saat urbanisasi berlangsung, jarak dari daerah asal ke daerah tujuan baru tidak lebih dari 350 km, alasan utama melakukan urbanisasi karena faktor pekerjaan (non militer) dan keluarga, jumlah anggota rumah tangga yang ikut urbanisasi memiliki jumlah yang cukup besar (1 hingga 9 orang) dan paling tidak jumlah anggota rumah tangga yang ikut pindah sebanyak 3 orang, jenjang pendidikan tertinggi lebih banyak di bawah SMA dan SMA, dan kelompok usia produktif cenderung lebih banyak melakukan urbanisasi. Metode klasifikasi pohon menghasilkan pohon optimal dengan ketepatan klasifikasi data *learning* sebesar 75,92%, sedangkan dari *testing* adalah sebesar 66,04% sehingga didapatkan faktor yang mempengaruhi urbanisasi adalah jarak lokasi pindah, jumlah anggota rumah tangga yang ikut pindah, jenjang pendidikan, lama waktu tinggal, dan alasan utama pindah.

DAFTAR PUSTAKA

- [1] BAPPENAS (2008), *Proyeksi Penduduk Indonesia (Indonesia Population Projection) 2005 – 2025* [Online]. Available: <http://bappenas.go.id/pencarian/global/?keyword=proyeksi+penduduk+indonesia&task=Cari>
- [2] A. Syaikat, "Faktor – Faktor Yang Menentukan Pilihan Daerah Tujuan Migrasi Penduduk Jawa Barat," Tesis Pasca Sarjana, Program Studi Kajian Kependudukan dan Ketenagakerjaan, Universitas Indonesia, Jakarta, Indonesia (1997).
- [3] H. Fadillah, "Faktor – Faktor Yang Mempengaruhi Proporsi Tujuan Migrasi Keluar Dari Kalimantan Selatan (Suatu Analisa Data SUPAS 1985)," Tesis Pasca Sarjana Program Studi Kajian Kependudukan dan Ketenagakerjaan, Universitas Indonesia. Universitas Indonesia, Jakarta, Indonesia (1993).
- [4] J. M. Vazquez, P. Panudulkitti dan Timofeev A. "Urbanization and The Poverty Level," International Studies Program Andrew Young School of Policy Studies Georgia State University, Georgia (2009).
- [5] D. E. Bloom, D. Canning D dan G. Fink (2008, Jan.). *Urbanization and The Wealth of Nations*. Harvard Initiative for Global Health [Online]. Available: http://www.hsph.harvard.edu/pgda/WorkingPapers/2008/PGDA_WP_30.pdf.
- [6] A. Andriyashin, "Financial Application of Classification and Regression Trees," Master Thesis, CASE-Center of Applied Statistics and Economics, Humboldt University, Berlin (2005).
- [7] Sumarmi, "Pendekatan Bagging CART Untuk Klasifikasi Anak Putus Sekolah di Jambi," Tesis Program Pasca Sarjana Jurusan Statistika, Institut Teknologi Sepuluh Nopember, Surabaya (2010).
- [8] I. B. Mantra, *Demografi Umum Edisi Kedua*, Yogyakarta : Pustaka Pelajar (2004).
- [9] L. Breiman, Friedman J.H, Olshen R.A dan Stone C.J., *Classification And Regression Tree*. New York: Chapman And Hall (1993).
- [10] R. J. Lewis, "An Introduction to Classification and Regression Tree (CART) Analysis," Presented at the 2000 Annual Meeting of the Society for Academic Emergency Medicine in San Fransisco, California.
- [11] R. A. Johnson dan G. K. Bhattacharyya, *Statistics Principles and Methods Third Edition*. Canada : John Wiley & Sons, Inc (1996).