

PENERAPAN METODE *QUEST* DALAM PEMBENTUKAN POHON KLASIFIKASI TINGKAT KEMISKINAN RUMAH TANGGA STUDI KASUS: RUMAH TANGGA DI KOTA PADANG

HUSNUL FIKRI, IZZATI RAHMI HG, DODI DEVIANTO

*Jurusan Matematika,
Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Andalas,
Kampus UNAND Limau Manis Padang, Indonesia,
email : husnulfikri@gmail.com izzatirahmihg@gmail.com ddvianto@yahoo.com*

Abstrak. Kemiskinan adalah permasalahan umum yang terjadi pada setiap negara maju dan berkembang, dimana kemiskinan adalah salah satu aspek yang dapat menghambat kesejahteraan keluarga dalam suatu rumah tangga. Pengeluaran per kapita rumah tangga adalah salah satu yang dapat dijadikan acuan dalam menentukan kriteria kemiskinan rumah tangga, dimana kriteria tersebut dibagi menjadi dua tingkatan berdasarkan ketetapan oleh Badan Pusat Statistik yaitu tidak miskin dengan pengeluaran per kapita rumah tangga dalam sebulan lebih dari Rp 350.610,00. Banyak faktor yang menyebabkan tingkatan kriteria kemiskinan berdasarkan pengeluaran per kapita rumah tangga. Analisa statistika yang dapat digunakan adalah metode *QUEST* (*Quick, Unbiased, Efficient, Statistical Tree*). Data yang diambil dari penelitian ini merupakan data sekunder dimana beberapa variabelnya diperoleh dari hasil SUSENAS 2015 dengan kriteria kemiskinan berdasarkan pengeluaran per kapita rumah tangga sebagai variabel tak bebas dan berbagai faktor yang menyebabkan kriteria kemiskinan sebagai variabel bebas. Setelah dilakukan analisa dengan metode *QUEST* diperoleh faktor yang berpengaruh dalam menentukan kriteria kemiskinan adalah kepemilikan *personal computer* dan banyaknya anggota rumah tangga.

Kata Kunci: Kriteria kemiskinan, pengeluaran perkapita rumah tangga, *QUEST*

1. PENDAHULUAN

Kemiskinan adalah permasalahan umum yang terjadi di setiap negara. Kemiskinan merupakan bagian dari krisis ekonomi, dimana krisis ekonomi merupakan salah satu aspek yang menghambat kesejahteraan keluarga di dalam suatu rumah tangga. Perlu diketahui faktor apa saja yang mempengaruhi tingkatan kemiskinan di rumah tangga agar terlihat bagaimana gambaran keluarga miskin dan keluarga tidak miskin di suatu rumah tangga.

Pada penelitian ini, kriteria kemiskinan adalah faktor yang akan diamati dalam menentukan sejahtera atau tidaknya suatu rumah tangga, kriteria kemiskinan dapat dilihat melalui pengeluaran per kapita rumah tangga. Pengeluaran perkapita rumah tangga perbulan adalah biaya yang dikeluarkan untuk konsumsi semua anggota rumah tangga selama satu bulan dibagi dengan banyaknya anggota rumah tangga dalam rumah tangga tersebut, maka dari itu pada penelitian ini dipilih pengeluaran per kapita rumah tangga selama satu bulan sebagai tolak ukur miskin atau

tidaknya suatu rumah tangga dan oleh karena itu dapat dikatakan bahwa tingkatan kemiskinan berdasarkan pengeluaran per kapita rumah tangga perbulan adalah variabel terikat atau *independent*.

Dalam hal ini faktor-faktor yang menyebabkan miskin atau tidak miskinnya suatu rumah tangga adalah faktor-faktor yang mempengaruhi kriteria kemiskinan, sehingga dapat dikatakan bahwa faktor-faktor yang mempengaruhi kriteria kemiskinan adalah variabel bebas atau *dependent*.

Analisa statistika yang digunakan untuk meneliti variabel-variabel bebas yang mempengaruhi variabel tak bebas dapat dilakukan teknik pengklasifikasian, Pengklasifikasian merupakan salah satu metode statistika untuk mengelompokkan atau mengklasifikasikan suatu data yang disusun secara sistematis. Masalah klasifikasi sering dijumpai dalam kehidupan sehari-hari diberbagai bidang kehidupan, baik dibidang ekonomi, sosial, pendidikan, maupun kesehatan [2].

Metode berstruktur pohon (*tree-structured method*) adalah salah satu metode dari pengklasifikasian suatu data dengan pendekatan non parametrik yang menyebar luas dalam berbagai bidang. Metode berstruktur pohon dapat dibagi menjadi dua kelompok yaitu kelompok pertama adalah metode yang menghasilkan pohon biner yaitu sebuah pohon yang disetiap simpul disekat menjadi dua simpul yang saling terpisah dan kelompok kedua adalah metode yang menghasilkan pohon non biner yaitu sebuah pohon yang setiap simpul disekat menjadi dua atau lebih simpul yang terpisah [4].

Metode yang akan diterapkan adalah metode *QUEST* pada data Survey Ekonomi Nasional rumah tangga di Kota Padang, dalam penelitian ini tingkat kemiskinan rumah tangga ditentukan berdasarkan pengeluaran per kapita per bulan rumah tangga pada data sebagai variabel tak bebas dan perilaku-perilaku secara umum suatu keluarga di rumah tangga yang menyebabkan tingkat kemiskinan sebagai variabel bebas, kemudian akan dilihat variabel bebas apa yang paling berpengaruh terhadap tingkat kemiskinan rumah tangga dengan metode *QUEST*. Penelitian ini bertujuan untuk mengaplikasikan metode *QUEST* dalam mengklasifikasikan data tingkatan kemiskinan rumah tangga dengan studi kasus di Kota Padang beserta faktor-faktor yang mempengaruhi tingkatan kemiskinan rumah tangga.

Dalam hal ini pemerintah Indonesia menetapkan pengertian kemiskinan oleh lembaga Badan Pusat Statistik (BPS). Lembaga tersebut mendefinisikan kemiskinan dengan membuat kriteria besarnya pengeluaran per kapita rumah tangga per bulan, kriteria statistik BPS tersebut adalah [3]:

- (1) Tidak miskin, yaitu pengeluaran per kapita rumah tangga dalam satu bulan lebih Rp 350.610,00
- (2) Hampir tidak miskin, yaitu pengeluaran per kapitarumah tangga dalam satu bulan berkisar antara Rp 280.488,00 dan Rp 350.60,00
- (3) Hampir miskin, yaitu pengeluaran per kapita rumah tangga dalam satu bulan berkisar antara Rp 233.740,00 dan Rp 280.488,00
- (4) Miskin, yaitu pengeluaran per perkapita rumah tangga dalam satu bulan kurang dari Rp 233.740,00

Kriteria yang ditetapkan tersebut berlandaskan kepada kebutuhan sehari-hari serta

kepemilikan dari masing-masing kepala keluarga, hal tersebut mencakup jenis dan luas lantai bangunan tempat tinggal, fasilitas kamar mandi dalam bangunan rumah, sumber penerangan rumah, sumber air minum, jumlah anggota keluarga, pendidikan tertinggi kepala rumah tangga, status kepemilikan rumah beserta fasilitas-fasilitas yang tersedia di dalam rumah.

2. Metode *QUEST*

QUEST (*Quick, Unbiased, Efficient Statistical Trees*) merupakan salah satu metode yang digunakan untuk membentuk pohon keputusan. *QUEST* merupakan algoritma pemisah yang menghasilkan pohon biner yang digunakan untuk klasifikasi. Algoritma *QUEST* memiliki kecepatan dalam hal komputasi (*Quick*), menghasilkan pemilihan variabel bebas yang tidak bias (*Unbiased*), dan efisien dalam pengolahan data (*Efficient*).

Algoritma *QUEST* dipisah menjadi tiga bagian, yaitu dengan uraian sebagai berikut [4] :

2.1. Algoritma Pemilihan Variabel Penyekat

Berikut adalah langkah-langkah pemilihan variabel penyekat :

- (1) Untuk setiap variabel X :
 - a. Jika X merupakan variabel kategorik, lakukan uji χ^2 untuk kebebasan antara variabel X dan variabel respon Y dan hitung nilai p dari pengujian tersebut.
 - b. Jika X merupakan variabel numerik, lakukan uji *ANOVA F* dan hitung nilai p dari pengujian tersebut.
- (2) Pilih variabel dengan nilai p terkecil atau bisa juga dilihat dari uji nilai statistik terbesar jika memiliki uji yang sama.
- (3) Bandingkan nilai p terkecil dengan taraf α/M_1 , dengan taraf nyata $\alpha \in (0, 1)$ dan M_1 adalah banyaknya variabel bebas.
 - a. Jika nilai p kurang dari α/M_1 , maka pilih variabel yang bersesuaian sebagai variabel penyekat. Teruskan ke langkah 5.
 - b. Jika nilai p lebih dari α/M_1 , teruskan ke langkah 4.
- (4) Untuk setiap variabel X yang numerik, maka hitung nilai p dari uji Levene untuk menguji kehomogenan ragam.
 - a. Pilih variabel dengan nilai p terkecil.
 - b. Bandingkan nilai p terkecil dari uji Levene dengan taraf $\alpha/M_1 + M_2$, dengan M_2 adalah banyaknya variabel bebas numerik.
 - c. Jika nilai p kurang dari $\alpha/M_1 + M_2$, maka pilih variabel yang bersesuaian sebagai variabel penyekat. Teruskan ke langkah 5.
 - d. Jika nilai p lebih dari $\alpha/M_1 + M_2$, maka variabel tersebut tidak dipilih menjadi variabel penyekat.
- (5) Misalkan X^* adalah variabel penyekat yang diperoleh dari langkah 3 atau 4, lanjutkan ke algoritma penentuan simpul penyekat.

2.2. Algoritma Penentuan Simpul Penyekat

Algoritma pemilihan simpul penyekat dapat dijelaskan sebagai berikut:

(1) Variabel Bebas Penyekat Kategorik

Jika variabel bebas X^* yang terpilih adalah nominal dan biner, maka simpul penyekat jelas. Namun jika variabel X^* yang terpilih adalah nominal dan lebih dari dua kategori, transformasikan variabel kategorik tersebut menjadi variabel numerik dengan langkah sebagai berikut:

- a. Transformasikan masing-masing nilai X^* ke vektor *dummy* L dimensi

$$\mathbf{v} = (v_1, v_2, \dots, v_L), \quad v_i = \begin{cases} 1, & x = b_i, \\ 0, & x \neq b_i \end{cases} \quad i = 1, 2, \dots, L$$

- b. Cari nilai tengah keseluruhan kelas k dari ν

$$\bar{\mathbf{v}} = \frac{\sum_{l=1}^L f_l \nu_l}{N_t}, \quad \bar{\mathbf{v}}^{(k)} = \frac{\sum_{l=1}^L n_{1l} \nu_l}{N_{k,t}},$$

dengan:

$\bar{\mathbf{v}}$: rata-rata untuk semua pengamatan pada simpul t ,

$\bar{\mathbf{v}}^{(k)}$: rata-rata pengamatan pada simpul t untuk kelompok ke k ,

f_l : jumlah pengamatan pada simpul t untuk ν_l ,

n_l : jumlah pengamatan pada simpul t kelompok ke k untuk ν_l ,

N_t : jumlah pengamatan pada simpul t ,

$N_{k,t}$: jumlah pengamatan pada simpul t untuk kelompok ke k .

- c. Tentukan matriks $L \times L$ berikut:

$$\mathbf{B} = \sum_{k=1}^K N_{k,t} (\bar{\mathbf{v}}^{(k)} - \bar{\mathbf{v}}) (\bar{\mathbf{v}}^{(k)} - \bar{\mathbf{v}}),$$

$$\mathbf{T} = \sum_{l=1}^L f_l (\nu_l - \bar{\mathbf{v}}) (\nu_l - \bar{\mathbf{v}}).$$

- d. Lakukan SVD (*Singular Value Decomposition*) pada \mathbf{T} untuk memperoleh nilai $\mathbf{T} = \mathbf{Q}\mathbf{D}\mathbf{Q}'$ dengan \mathbf{Q} adalah matriks orthogonal, $\mathbf{D} = \text{diag}(d_1, d_2, \dots, d_L)$ dengan $d_1 \geq d_2 \geq \dots \geq d_L \geq 0$. Kemudian tentukan $D^{-1/2}$ dengan memisalkan $D^{-1/2} = \text{diag}(d_1^*, d_2^*, \dots, d_L^*)$ dimana:

$$d_l^* = \begin{cases} d_l^{1/2}, & d_l > 0, \\ 0, & \text{selainnya.} \end{cases}$$

- e. Lakukan SVD dari $\mathbf{D}^{-1/2}\mathbf{Q}'\mathbf{B}\mathbf{Q}\mathbf{D}^{-1/2}$ untuk memperoleh vektor eigen α yang berhubungan dengan nilai eigen terbesar.

- f. Tentukan koordinat diskriminan terbesar dari ν yaitu :

$$\xi = \alpha' D^{1/2} Q' D'.$$

(2) Variabel Penyekat Numerik

Variabel tidak bebas Y terdiri dari dua kategori yaitu dimisalkan dengan 0 dan 1 dan variabel bebas X terpilih berupa numerik. Langkah penentuan variabel bebas penyekat adalah sebagai berikut.

- a. Definisikan \bar{x}_0 dan s_0^2 adalah nilai tengah dan ragam X^* dari pengamatan dengan respon 0, sedangkan \bar{x}_1 dan s_1^2 adalah nilai tengah dan ragam X^* dari pengamatan dengan respon 1. Misalkan $P(k|t) = N_{k,t}/N_k$ merupakan peluang dari masing-masing kategori variabel respon, dengan $N_{k,t}$ adalah jumlah data pada simpul t untuk respon k dan N_k adalah jumlah data pada simpul awal untuk respon k .
- b. Jika $\min(s_0^2, s_1^2) = 0$ maka urutkan kedua kelas dengan urutan ragam yang meningkat kemudian notasikan ragam tersebut, misalkan $s_0^2 = 0$ maka $s_0^2 \geq s_1^2$ dan disesuaikan dengan rata-ratanya yaitu \bar{x}_0 untuk ragam pada kategorik 0 dan \bar{x}_1 untuk ragam pada kategorik 1.
- c. Jika $\min(s_0^2, s_1^2) \neq 0$ maka analisis diskriminan kuadratik diterapkan untuk menghitung simpul penyekat d . Analisis diskriminan kuadratik menyekat sumbu X^* menjadi tiga interval yaitu (∞, d_1) , (d_1, d_2) dan (d_2, ∞) dimana d merupakan akar-akar dari persamaan:

$$P(0 | t) s_0^{(-1)} \varphi\left(\frac{x - \bar{x}_0}{s_0}\right) = P(1 | t) s_1^{(-1)} \varphi\left(\frac{x - \bar{x}_1}{s_1}\right) \quad (2.1)$$

Solusi tersebut dapat ditentukan dengan menentukan akar persamaan kuadrat $ax^2 + bx + c = 0$ dengan

$$\begin{aligned} a &= s_0^2 - s_1^2, \\ b &= 2(\bar{x}_0 s_1^2 - \bar{x}_1 s_0^2), \\ c &= (\bar{x}_1 s_0)^2 - (\bar{x}_0 s_1)^2 + 2s_0^2 s_1^2 \ln \left\{ \frac{P(0 | t)}{P(1 | t)} \right\}. \end{aligned}$$

- d. Simpul disekat pada titik $X^* = d$ dimana d didefinisikan sebagai berikut:

- Jika $\bar{x}_0 < \bar{x}_1$ maka $d = \bar{x}_0$
- Jika $a = 0$ maka:

$$d = \begin{cases} \frac{\bar{x}_0 + \bar{x}_1}{2} - (\bar{x}_0 + \bar{x}_1^{(-1)}) s_0^2 \ln \left\{ \frac{P(0 | t)}{P(1 | t)} \right\}, & \bar{x}_0 \neq \bar{x}_1 \\ \bar{x}_0, & \bar{x}_0 = \bar{x}_1. \end{cases}$$

- Jika $a \neq 0$ maka :

- * Jika $b^2 - 4ac < 0$ maka $d = \frac{1}{2}(\bar{x}_0 + \bar{x}_1)$

- * Jika $b^2 - 4ac \geq 0$ maka d adalah akar dari $\frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$ yang lebih mendekati nilai \bar{x}_0 dengan syarat menghasilkan simpul tak kosong.

2.3. Algoritma Pemberhentian Pembentukan Pohon

Proses penyekatan dilakukan sampai simpul tidak bisa disekat lagi dengan aturan penghentian proses pembentukan pohon yang ditentukan, antara lain sebagai berikut:

- a. Jika kedalaman pohon pada saat itu telah mencapai titik maksimum yang ditetapkan, maka proses penyekatan pohon dihentikan.
- b. Jika penyekat dari simpul menghasilkan simpul anak yang ukuran simpulnya kurang dari nilai ukuran simpul anak minimum yang ditetapkan, penyekatan akan dihentikan.

3. DATA DAN METODE PENELITIAN

3.1. Data

Data yang digunakan pada penelitian ini merupakan data sekunder yang diambil dari hasil Survey Sosial Ekonomi Nasional (SUSENAS) tahun 2015 Kota Padang. Variabel yang digunakan dalam penelitian ini terdiri dari:

- (1) Variabel tak bebas (Y) yakni tingkatan kemiskinan berdasarkan pengeluaran per kapita per bulan rumah tangga, dimana tingkatan kemiskinan berdasarkan pengeluaran per kapita per bulan rumah tangga terbagi atas:
 - 1 : Rumah tangga miskin,
 - 2 : Rumah tangga tidak miskin.
- (2) Variabel bebas (X) terdiri atas 13 variabel yaitu jenis kelamin kepala keluarga (X_1), status perkawinan kepala keluarga, (X_2), umur kepala keluarga (X_3), status pekerjaan kepala keluarga (X_4), banyaknya anggota rumah tangga (X_5), status kepemilikan rumah (X_6), cara memperoleh air minum (X_7), sumber air minum (X_8), sumber penerangan utama (X_9), jenis bahan bakar dalam memasak (X_{10}), menerima beras miskin (X_{11}), kepemilikan telepon (X_{12}), kepemilikan personal computer (X_{13})

3.2. Metode Analisis

- (1) Akan dilakukan uji kebebasan antara variabel bebas dan variabel tak bebas dengan uji khi kuadrat untuk variabel kategorik dan uji kesamaan nilai tengah dengan uji *ANOVA F* untuk variabel numerik.
- (2) Sebagai tahap akhir, akan dilakukan analisis data dengan metode *QUEST* berdasarkan algoritma pembentukan pohon klasifikasi *QUEST* dengan bantuan software *IBM SPSS statistic 24*.

4. HASIL DAN PEMBAHASAN

4.1. Uji Hipotesis Kebebasan dan Uji Hipotesis Kesamaan Nilai Tengah

- (1) Untuk variabel bebas kategorik
Untuk menguji hipotesis kebebasan antara variabel bebas kategorik dan variabel

tak bebas, diterapkan uji khi kuadrat, dengan hipotesis sebagai berikut

$$H_0 : X \text{ dan } Y \text{ saling bebas,}$$

$$H_1 : X \text{ dan } Y \text{ tidak saling bebas,}$$

Statistik uji yang digunakan adalah

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^c \left[\frac{(o_{ij} - E_{ij})^2}{E_{ij}} \right],$$

dengan r adalah baris dan c adalah kolom

$$E_{ij} = \frac{n_{.i}n_{.j}}{n}.$$

H_0 ditolak jika nilai dari $\chi^2 > \chi_{hitung}^2$ dengan χ^2 adalah variabel acak khi-kuadrat dengan $db = (r - 1)(c - 1)$

(2) Untuk variabel bebas numerik

Untuk menguji kesamaan nilai tengah antara variabel bebas numerik dan variabel tak bebas, diterapkan uji ANOVA F , dengan hipotesis sebagai berikut

$$H_0 : \mu_1 = \mu_2,$$

$$H_1 : \mu_1 \neq \mu_2.$$

Statistik uji yang digunakan adalah

$$F = \frac{s_1^2}{s_2^2},$$

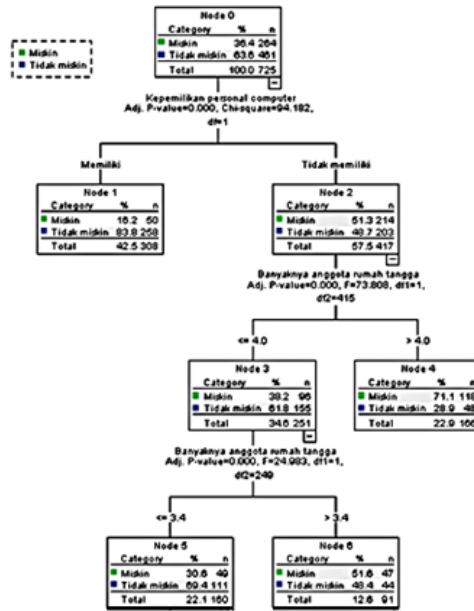
dengan

$$s_1^2 = \left(\frac{\sum_{k=1}^K \frac{T_{k.}^2}{n_K} - \frac{T_{..}^2}{N}}{K - 1} \right),$$

$$s_2^2 = \frac{\left(\sum_{k=1}^K \sum_{i=1}^{n_K} x_{ki}^2 - \frac{T_{..}^2}{N} \right) - \left(\sum_{k=1}^K \frac{T_{k.}^2}{n_K} \right)}{K - 1}.$$

Jika $F_{hitung} > F_{(\infty,1,n-2)}$ maka tolak H_0 yang berarti bahwa variabel memiliki nilai tengah yang tidak sama sehingga variabel tersebut berpengaruh terhadap variabel Y .

Bila digunakan $\alpha = 0,05$ didapat $\frac{\alpha}{M_1} = \frac{0,05}{13} = 0,0038$. Dengan menggunakan bantuan *software IBM SPSS Statistic* terdapat beberapa variabel bebas yang nilai $p > 0,0038$ sehingga variabel bebas tersebut tidak berpengaruh terhadap variabel tak bebas, variabel bebas tersebut antara lain jenis kelamin (X_1), umur kepala keluarga (X_3) dan sumber penerangan utama (X_9).



Gambar 1. Pohon Klasifikasi tingkatan kemiskinan rumah tangga di Kota Padang dengan Metode *QUEST*

4.2. Pohon Klasifikasi

Pada Gambar 1 diberikan pohon kualifikasi tingkatan kemiskinan rumah tangga di Kota Padang dengan Metode *QUEST*. Terdapat dua variabel yang berpengaruh terhadap tingkatan kemiskinan berdasarkan pengeluaran per kapita rumah tangga di Kota Padang dengan metode *QUEST*. Kedua variabel tersebut adalah kepemilikan *personal computer* dan banyaknya anggota rumah tangga, karakteristik yang diperoleh dari hasil dengan metode *QUEST* adalah:

1. Kelas pertama yaitu rumah tangga yang memiliki *personal computer*. Rumah tangga dengan karakteristik ini adalah rumah tangga yang tidak miskin.
2. Kelas kedua yaitu rumah tangga yang tidak memiliki *personal computer* dengan anggota rumah tangga kurang atau sama dari 4 orang.
3. Kelas selanjutnya yaitu rumah tangga yang tidak memiliki *personal computer* dengan banyak anggota rumah tangga 4 orang. Rumah tangga dengan karakteristik ini adalah rumah tangga yang miskin.

5. PENUTUP

Setelah dilakukan pengklasifikasian dengan metode *QUEST* faktor yang paling berpengaruh terhadap tingkatan kemiskinan adalah kepemilikan *personal computer* dan banyaknya anggota rumah tangga.

Pada penulisan ini, saran untuk penulisan selanjutnya adalah:

- (1) Pada penelitian ini yang menjadi faktor yang paling signifikan dalam menentukan kriteria kemiskinan adalah kepemilikan *personal computer*, dalam hal ini, ada faktor yang mungkin hilang jika dikaitkan dengan kepemilikan *personal computer*. Faktor tersebut tidak dimuat didalam variabel-variabel yang ada pada data.
- (2) Memperjelas secara rinci variabel terkait dengan jenis pekerjaan sehingga diketahui tingkat kebutuhannya serta penambahan variabel-variabel lain yang penting terkait dengan faktor yang mempengaruhi kemiskinan rumah tangga.

Daftar Pustaka

- [1] Bain, Lee J dan Max Engelhardt. 1992. *Introduction to Probability and Mathematical Statistics*, Second Edition. Duxbury Press, California.
- [2] Breiman, et.al. 1984. *Classification and Regression Tree*. Chapman & Hall, New York.
- [3] Badan Pusat Statistik. 2015. *Survei Sosial Ekonomi Nasional (SUSENAS) Modul Sosial Budaya dan Pendidikan (Tahunan)*.
<https://sirusa.bps.go.id/sirusa/index.php/dasar/pdf?kd=1581& th=2015>, diakses pada 20 November 2016.
- [4] Loh W-Y, Shin Y-S. 1997. *Split Selection Methods for Classification Trees*. Statistic Sinica.