

PENERAPAN METODE *QUICK, UNBIASED, EFFICIENT STATISTICAL TREES (QUEST)* UNTUK MENENTUKAN FAKTOR-FAKTOR YANG MEMPENGARUHI PENYAKIT DIARE PADA BALITA DI INDONESIA

Vara Diba Rizki¹, Yudi Setyawan²

^{1,2}Jurusan Statistika, FST, Institut Sains & Teknologi AKPRIND Yogyakarta
E-mail: varadibarizki@yahoo.co.id

Abstract. Classification is a systematic grouping of objects into a particular class or pattern based on similarity of features. One of the statistical methods that can be used to classify objects is the *QUEST* method. The *QUEST* method (Quick, Unbiased, Efficient, Statistical Trees) is a tree-classification method that generates binary tree which is fast in computation, produce unbiased variables, and use efficient algorithms. This study discusses the classification by using the *QUEST* method in determining factors that affect diarrhea diseases suffered by under 5 year age in Indonesia. Data of Indonesia Demographic and Health Survey (IDHS 2012) shows the percentage of under 5 years old of age children affected by diarrhea in Indonesia by 14% or as many as 2,342 people from 16,380 samples, while 86% or 14,038 others are children without diarrhea. The binary tree classification procedure using the *QUEST* method is: selecting the partition from each independent variable, determining the procedure of partition point and terminating of the tree formation. In these procedure, the factors that influence diarrhea disease among children are age, gender, toilet facility, residence area, mother education, and wealth quantity. The most influence variable for under 5 year children diarrhea is age variable. Classification accuracy using *QUEST* method is 91,1%.

Keywords: *QUEST* Method, Children of diarrhea.

Abstrak. Pengklasifikasian merupakan pengelompokan secara sistematis terhadap suatu objek ke dalam golongan atau pola-pola tertentu berdasarkan kesamaan ciri. Salah satu metode statistika yang dapat digunakan untuk mengklasifikasikan objek adalah metode *QUEST*. Metode *QUEST* (Quick, Unbiased, Efficient, Statistical Trees) merupakan salah satu metode klasifikasi berstruktur pohon yang menghasilkan pohon biner dan bersifat cepat dalam komputasi, menghasilkan pemilihan peubah bebas yang takbias, serta memiliki algoritma yang efisien. Penelitian ini membahas pengklasifikasian dengan menggunakan metode *QUEST* dalam menentukan faktor-faktor yang mempengaruhi penyakit diare pada balita di Indonesia. Data Survei Demografi dan Kesehatan Indonesia (SDKI) 2012 menunjukkan persentase balita terkena diare di Indonesia sebesar 14% atau sebanyak 2.342 orang dari 16.380 orang sampel, sedangkan 86% atau 14.038 orang lainnya merupakan balita tidak terkena diare. Prosedur klasifikasi berstruktur pohon biner dengan metode *QUEST* yaitu: proses memilih penyekat dari tiap peubah bebas, menentukan titik penyekat dan penghentian pembentukan pohon. Pada prosedur tersebut menghasilkan faktor-faktor yang mempengaruhi penyakit diare pada balita diantaranya umur anak, jenis kelamin, fasilitas kakus, daerah tempat tinggal, pendidikan ibu, dan kuantil kekayaan. Peubah yang paling berpengaruh pada balita diare adalah umur anak. Ketepatan klasifikasi menggunakan metode *QUEST* sebesar 91,1%.

Kata Kunci: Metode *QUEST*, Diare pada balita.

1. Pendahuluan

Pengklasifikasian merupakan pengelompokan secara sistematis terhadap suatu objek ke dalam golongan atau pola-pola tertentu berdasarkan kesamaan ciri. Salah satu metode statistika yang dapat digunakan untuk mengklasifikasikan objek adalah metode *QUEST* (Quick, Unbiased, Efficient, Statistical Trees). Metode ini merupakan salah satu metode klasifikasi berstruktur pohon yang menghasilkan pohon biner dan bersifat cepat dalam komputasi, menghasilkan pemilihan peubah bebas yang takbias, serta memiliki algoritma yang efisien. Metode *QUEST* digunakan ketika data pada peubah respon bertipe kualitatif dan peubah bebas bertipe data kualitatif atau kuantitatif.

Penggunaan metode *QUEST* telah banyak digunakan di berbagai bidang di antaranya bidang kesehatan, pendidikan, sosial, industri, dan perbankan. Pengklasifikasian pada bidang kesehatan dapat diterapkan untuk mengetahui faktor-faktor yang berhubungan dengan kejadian suatu penyakit. Salah satu penyakit yang banyak terjadi pada bayi dan balita di Indonesia adalah penyakit diare. Menurut data *World Health Organization* (2009), diare adalah penyebab kematian kedua pada anak di bawah 5 tahun. Setiap anak di Indonesia rata-rata mengalami episode diare sebanyak 1,6-2 kali per tahun (DepKes RI, 2009).

Penelitian ini akan membahas metode klasifikasi berstruktur pohon menggunakan algoritma *QUEST* yang bertujuan untuk mengetahui prosedur klasifikasi berstruktur pohon biner dengan metode *QUEST*, menerapkan metode *QUEST* dalam mengidentifikasi faktor-faktor apa yang mempengaruhi penyakit diare pada balita di Indonesia serta mengetahui faktor yang paling berpengaruh atau menjadi faktor penciri penyakit diare pada balita di Indonesia.

Pengklasifikasian dengan metode *QUEST* pada penelitian sebelumnya telah dilakukan oleh Handayani P., tahun 2013 (*STIKes WDH Repository*) dalam menggambarkan karakteristik kejadian diare pada balita. Alwi W., dkk., tahun 2014 (*ANDALAS Repository*) yang bertujuan untuk mengetahui faktor-faktor yang mempengaruhi seseorang menderita penyakit kencing manis. Septiyani T., tahun 2016 (*UNS Repository*) yang bertujuan untuk mengetahui faktor-faktor yang mempengaruhi pasien penderita stroke.

2. Metode Penelitian

2.1 Sumber Data

Data yang digunakan dalam penelitian ini adalah data sekunder, yaitu data balita diare di Indonesia berdasarkan Survei Demografi dan Kesehatan Indonesia (SDKI) tahun 2012.

2.2 Variabel Penelitian

Variabel dalam penelitian ini terdiri atas variabel dependen (*Y*) dan variabel independen (*X*). Variabel dependen (*Y*) adalah status balita diare, dan variabel independen (*X*) adalah umur anak, jenis kelamin, sumber air minum, fasilitas kakus, daerah tempat tinggal, pendidikan ibu, dan kuantil kekayaan.

2.3 Metode Analisis Data

2.3.1 Metode *QUEST*

Metode *QUEST* (*Quick, Unbiased, Efficient Statistical Trees*) merupakan pengembangan dari *FACT* (*Factor Analysis Classification Trees*) yang memiliki kecepatan komputasi yang tinggi dan suatu metode pohon klasifikasi yang menghasilkan pohon biner (Loh dan Shih, 1997). Algoritma pembentukan pohon klasifikasi ini merupakan modifikasi dari analisis diskriminan kuadrat. Algoritma *QUEST* dibagi menjadi tiga bagian yaitu, algoritma pemilihan peubah penyekat, algoritma penentuan simpul penyekat, dan algoritma penghentian pembentukan pohon.

1. Algoritma Pemilihan Peubah Penyekat

Pemilihan peubah penyekat suatu simpul dilakukan dengan menggunakan uji dependen independensi dan kesamaan rata-rata. Peubah bebas yang memiliki P_{value} terkecil menjadi peubah penyekat awal pada pohon klasifikasi. Setiap peubah bebas memiliki kesempatan untuk terpilih sebagai peubah penyekat, meskipun peubah tersebut telah terpilih sebagai peubah penyekat untuk simpul sebelumnya. Apabila peubah bebas berskala nominal pemilihan peubah penyekat menggunakan uji dependen independensi dengan menggunakan Statistik *chi-square* yang dapat diperoleh melalui persamaan berikut:

$$\chi^2_{hitung} = \sum_{i=1}^b \sum_{j=1}^k \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \quad (1)$$

Jika $\chi^2_{hitung} > \chi^2_{tabel}$ atau $P_{value} < \alpha$ maka H_0 ditolak dan disimpulkan bahwa terdapat hubungan antara peubah bebas dengan peubah respon.

Apabila peubah bebas berskala ordinal atau kontinu pemilihan peubah penyekat menggunakan uji kesamaan rata-rata. Misalkan x_{ij} merupakan pengamatan ke- j dari kelompok ke- i , maka dapat disajikan struktur data seperti pada Tabel 1.

Tabel 1. k sampel acak

	Kelompok				
	1	2	...	K	
	x_{11}	x_{21}	...	x_{k1}	
	x_{12}	x_{22}	...	x_{k2}	
	\vdots	\vdots		\vdots	
	x_{1n_1}	x_{1n_2}	...	x_{kn_k}	
Jumlah	T_1	T_2	...	T_k	$T_{..}$
Rataan	\bar{x}_1	\bar{x}_2	...	\bar{x}_k	$\bar{x}_{..}$

Bila μ_i adalah rata-rata dari kelompok ke- i ($i = 1, 2, \dots, k$) maka hipotesis yang digunakan dalam uji ini adalah:

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_k$$

$$H_1 : \text{paling sedikit ada dua rataan tidak sama.}$$

Berikut Tabel 2 yang digunakan untuk mendapatkan statistik hitung yaitu f_h .

Tabel 2. Analisis variansi untuk klasifikasi

Sumber variasi	Jumlah kuadrat	Derajat Bebas	Rataan Kuadrat	f hitung
Kelompok	JKK	$k - 1$	$s_1^2 = \frac{JKK}{k - 1}$	$f_h = \frac{s_1^2}{s^2}$
Galat	JKG	$n - k$	$s^2 = \frac{JKG}{n - k}$	
Total	JKT	$n - 1$		

dengan :

$$JKT = \sum_{i=1}^k \sum_{j=1}^{n_i} x_{ij}^2 - \frac{T_{..}^2}{n} \quad (2)$$

$$JKK = \sum_{i=1}^k \frac{T_i^2}{n_i} - \frac{T_{..}^2}{n} \quad (3)$$

$$JKG = JKT - JKK$$

$$i = 1, 2, \dots, k, \quad j = 1, 2, \dots, n_i.$$

Dalam pengambilan keputusan, H_0 ditolak pada taraf α jika nilai $f_h > f_{\alpha[(k-1), (n-k)]}$.

Sebelum dilakukan uji kesamaan rata-rata dilakukan terlebih dahulu uji homogenitas ragam pada peubah bebas berskala ordinal atau kontinu. Hipotesis yang digunakan dalam uji ini adalah:

$$H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2$$

$$H_1 : \text{paling sedikit ada dua ragam tidak sama}$$

Statistik uji diperoleh dengan mentransformasikan data menjadi simpangan mutlak terhadap rata-rata data, yaitu $y_{ij} = |x_{ij} - \bar{x}_i|$,

dengan :

$$i = 1, 2, \dots, k \quad j = 1, 2, \dots, n_i$$

y_{ij} : hasil transformasi data pengamatan ke- j dari kelompok ke- i

\bar{x}_i : rata-rata sampel kelompok ke- i

x_{ij} : pengamatan ke- j dari kelompok ke- i .

Lakukan statistik uji kesamaan rata-rata pada data yang telah ditransformasi untuk mendapatkan nilai statistik uji homogenitas ragam.

Dalam pengambilan keputusan, H_0 ditolak pada taraf α jika nilai $f_h > f_{\alpha}[(k-1), (n-k)]$.

2. Algoritma Penentuan Titik Penyekat

Setelah peubah penyekat terpilih, lakukan algoritma penentuan titik penyekat untuk menentukan titik penyekat. Langkah-langkah penentuan titik penyekat dapat diuraikan sebagai berikut:

Misalkan peubah respon terdiri dari dua kategori dan peubah bebas penyekat X^* yang terpilih berupa peubah ordinal atau kontinu. Langkah - langkah penentuan peubah bebas penyekat sebagai berikut:

1. Definisikan \bar{x}_0 dan s_0^2 adalah rata-rata dan ragam X^* dari pengamatan dengan respon 0, sedangkan \bar{x}_1 dan s_1^2 adalah rata-rata dan ragam X^* dari pengamatan dengan respon 1. Misalkan $P(k|t) = N_{k,t}/N_k$ merupakan peluang dari masing-masing kategori peubah 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 .
2. Tentukan solusi dari persamaan:

$$P(0|t)s_0^{-1}\varphi\left(\frac{x - \bar{x}_0}{s_0}\right) = P(1|t)s_1^{-1}\varphi\left(\frac{x - \bar{x}_1}{s_1}\right) \quad (4)$$

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

$$a = s_0^2 - s_1^2$$

$$b = 2((\bar{x}_0 s_1^2 - \bar{x}_1 s_0^2)$$

$$c = (\bar{x}_1 s_0)^2 - (\bar{x}_0 s_1)^2 + 2s_0^2 s_1^2 \ln \left\{ \frac{P(0|t)s_1^2}{P(1|t)s_0^2} \right\}$$

3. Simpul disekat pada titik $X^* = d$, di mana d didefinisikan sebagai berikut:

a. Jika $\bar{x}_0 < \bar{x}_1$, maka $d = \bar{x}_0$

b. Jika $a = 0$, maka:

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

c. 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 $\frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$ lebih mendekati nilai \bar{x}_0 , dengan syarat menghasilkan dua simpul tak-kosong.

Misalkan peubah respon terdiri dari dua kategori dan peubah bebas penyekat X^* yang terpilih berupa peubah nominal. Berikut adalah langkah-langkah penentuan titik penyekat:

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

$$v = (v_1, v_2, \dots, v_L)'$$

$$\text{dengan } v_l = \begin{cases} 1 & , \text{ untuk } x = b_l \\ 0 & , \text{ untuk } x \neq b_l \end{cases} , l = 1, 2, \dots, L$$

2. Cari nilai rata-rata untuk X^* yang telah ditransformasi:

$$\bar{v} = \frac{\sum_{l=1}^L f_l v_l}{N_t} \quad (5)$$

$$\bar{v}^{(k)} = \frac{\sum_{l=1}^L n_{l,t} v_l}{N_{k,t}} \quad (6)$$

dengan :

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

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

f_l : jumlah pengamatan pada simpul t untuk v_l

$n_{l,t}$: jumlah pengamatan pada simpul t kelompok ke- k untuk v_l

N_t : jumlah pengamatan pada simpul t

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

3. Tentukan matriks $L \times L$ berikut :

$$B = \sum_{k=1}^K N_{k,t} (\bar{v}^{(k)} - \bar{v})(\bar{v}^{(k)} - \bar{v}) \quad (7)$$

$$T = \sum_{l=1}^L f_l (v_l - \bar{v})(v_l - \bar{v})' \quad (8)$$

4. Lakukan SVD (*Singular Value Decomposition*) terhadap matriks T . Menentukan SVD meliputi langkah-langkah menentukan nilai-nilai eigen dan vector-vector eigen dari matriks TT' atau $T'T$. Dekomposisi nilai singular matriks riil T $m \times n$ adalah faktorisasi $T = QDQ'$, dengan Q adalah matriks orthogonal $L \times L$ yang tersusun atas vektor-vektor terurut dari $T'T$, $D = \text{diag}(d_1, \dots, d_L)$ dengan $d_1 \geq d_2 \geq \dots \geq d_L \geq 0$, matriks diagonal $m \times n$ bernilai riil tak negatif yang disebut nilai-nilai singular.

5. Tentukan $D^{-\frac{1}{2}} = \text{diag}(d_1^*, \dots, d_L^*)$,

$$\text{dengan } d_i^* = \begin{cases} d_i^{-1/2} & , \text{ untuk } d_i > 0 \\ 0 & , \text{ untuk lainnya.} \end{cases}$$

6. Lakukan SVD dari $D^{-\frac{1}{2}} Q' B Q D^{-\frac{1}{2}}$, tentukan vektor eigen a yang merupakan vektor eigen yang berpadanan dengan nilai eigen terbesar.

7. Tentukan koordinat diskriminan terbesar dari v pada langkah 1, yaitu :

$$\xi = a' D^{-\frac{1}{2}} Q' v \quad (9)$$

Setelah diperoleh nilai ξ selanjutnya dilakukan analisis diskriminan kuadratik untuk menentukan titik penyekat.

3. Algoritma Penghentian Pembentukan Pohon

Proses pemilihan simpul secara rekursif akan berhenti apabila salah satu dari aturan penghentian tercapai. Ada empat aturan penghentian pembentukan pohon yaitu:

- 1) Tidak ada lagi peubah bebas yang signifikan menunjukkan perbedaan terhadap peubah respon.
- 2) Jika kedalaman pohon pada saat itu telah mencapai batas titik maksimum yang ditetapkan, maka proses pertumbuhan akan berhenti.

- 3) Jika penyekat dari simpul menghasilkan *parent node* dan *child node* yang ukuran simpulnya kurang dari nilai ukuran *parent node* dan *child node* minimum yang ditetapkan, penyekatan akan dihentikan.

2.3.2 Ketepatan Klasifikasi

Salah satu ukuran kebaikan model dalam regresi logistik yaitu jika memiliki peluang salah klasifikasi yang minimal (Hosmer & Lemeshow, 1989). Ketepatan klasifikasi (*correct classification*) terhadap amatan pada regresi logistik biner harus menentukan nilai *cutpoint* (c) dan dibandingkan dengan peluang dugaan $\pi(x)$. Jika $\pi(x)$ lebih besar atau sama dengan c maka nilai dugaan termasuk pada respon $y=1$ dan selain itu $y=0$. Seperti halnya pada metode regresi logistik biner, ketepatan dan kesalahan klasifikasi juga dapat diketahui dari tabel klasifikasi metode CHAID. Ketepatan klasifikasi (*correct classification*) terdiri atas *specificity* dan *sensitivity*. *Specificity* dan *sensitivity* dapat dihitung menggunakan tabel berikut:

Tabel 3. Klasifikasi respon

Amatan	Dugaan	
	1	0
1	a	b
0	c	d

Specificity atau ketepatan klasifikasi dalam menduga kejadian bahwa respon tidak memiliki kriteria yang diharapkan yaitu pada ($y=0$) sebesar $\left(\frac{d}{c+d}\right) \cdot 100\%$, untuk mengevaluasi ketepatan klasifikasi dalam menduga kejadian bahwa respon memiliki kriteria yang diharapkan yaitu ($y=1$) atau disebut juga *sensitivity* yang nilainya sebesar $\left(\frac{a}{a+b}\right) \cdot 100\%$, sedangkan ketepatan klasifikasi yaitu ketepatan klasifikasi dalam menduga kejadian secara tepat dapat diduga oleh model yang nilainya $\left(\frac{a+d}{(a+b+c+d)}\right) \cdot 100\%$.

3 Hasil Dan Pembahasan

3.1 Analisis Deskriptif Balita Terkena Diare Secara Umum

Persentase balita terkena diare berdasarkan data Survei Demografi dan Kesehatan Indonesia (SDKI) sebesar 14% atau sebanyak 2.342 orang dari 16.380 orang sampel, sedangkan 86% atau 14.038 orang lainnya merupakan balita tidak terkena diare.

3.2 Pembentukan Pohon Klasifikasi QUEST Penyakit Diare pada Balita Di Indonesia

Pembentukan pohon klasifikasi biner menggunakan algoritma *QUEST* dapat dibagi menjadi tiga bagian yaitu algoritma pemilihan peubah penyekat, algoritma pemilihan titik penyekat, dan algoritma penghentian pembentukan pohon. Peubah yang digunakan dalam penelitian ini adalah peubah respon Y yang terdiri dari dua kategori yaitu $j = 1$ dan 2, kategori 1 menyatakan balita diare dan kategori 2 menyatakan balita tidak diare. Sementara untuk peubah bebas terdiri dari peubah umur anak (X_1), jenis kelamin (X_2), sumber air minum (X_3), fasilitas kakus (X_4), daerah tempat tinggal (X_5), pendidikan ibu (X_6), dan kuantil kekayaan (X_7). Berikut dijelaskan pembentukan pohon klasifikasi biner menggunakan algoritma *QUEST* pada kasus penyakit diare pada balita di Indonesia.

1. Algoritma Pemilihan Peubah Penyekat

Pemilihan peubah penyekat suatu simpul dilakukan dengan menggunakan uji dependen independensi untuk peubah berskala nominal dan uji kesamaan rata-rata untuk peubah berskala ordinal. Peubah bebas yang memiliki P_{value} terkecil menjadi peubah penyekat awal pada pohon klasifikasi. Sebagai ilustrasi, berikut ini dilakukan secara manual perhitungan kesamaan rata-

rata untuk peubah X_1 berskala ordinal dan uji dependen independensi untuk peubah X_2 berskala nominal.

Pada peubah X_1 dilakukan terlebih dahulu uji homogenitas ragam. Hipotesis yang digunakan dalam uji ini adalah:

$$H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2$$

H_1 : paling sedikit ada dua ragam tidak sama

Statistik uji diperoleh dengan mentransformasikan data menjadi simpangan mutlaknya terhadap rata-rata data, yaitu $y_{ij} = |x_{ij} - \bar{x}_i|$. Untuk peubah X_1 , misalkan T_1 , merupakan jumlah pengamatan setelah transformasi untuk peubah X_1 pada balita terkena diare, T_2 , merupakan jumlah pengamatan setelah transformasi untuk peubah X_1 pada balita tidak terkena diare, T_3 dan merupakan jumlah pengamatan setelah transformasi peubah X_1 untuk seluruh data sehingga diperoleh $T_1 = 2729,93$ $T_2 = 18836,53$, $T_3 = 21566,46$, sehingga:

Tabel 4. Analisis variansi setelah dilakukan transformasi

Sumber variasi	Jumlah kuadrat	Derajat Bebas	Rataan Kuadrat	f hitung
Kelompok	62,305	1	62,305	85,213
Galat	11975,058	16378	0,731	
Total	12037,362	16379		

Dengan bantuan Ms.Excel diperoleh $P_{value} = 0,000 < \frac{\alpha}{M_1+M_2} = 0,005$ dalam hal ini H_0

ditolak dan disimpulkan bahwa ragam kedua kelompok tidak sama atau heterogen sehingga dapat dilakukan uji kesamaan rata-rata peubah X_1 .

Uji kesamaan rata-rata peubah status diare dan peubah umur anak.

Uji Hipotesis:

1. H_0 : peubah status diare dan peubah umur anak adalah independen
 H_1 : peubah status diare dan peubah umur anak adalah dependen
2. Tingkat signifikansi $\alpha = 0,05$
3. Kriteria uji : H_0 ditolak jika $f_h > f_{\alpha}[(k-1),(n-k)]$ atau $P_{value} < 0,05/7$
4. Statistik uji:

Pada peubah X_1 , jika $j = 1,2$ dengan 1 menyatakan kelas balita diare dan 2 menyatakan kelas balita tidak diare. Untuk peubah X_1 , misalkan T_1 , merupakan jumlah pengamatan peubah X_1 pada balita diare, T_2 , merupakan jumlah pengamatan peubah X_1 pada balita tidak diare dan T_3 merupakan jumlah pengamatan peubah X_1 untuk seluruh data, diperoleh:

$$\begin{aligned} T_1 &= 8196 & n_1 &= 2342 \\ T_2 &= 54967 & n_2 &= 14038 \\ T_3 &= 63163 & n &= 16380 \end{aligned}$$

Sehingga diperoleh nilai F_{hitung} seperti pada tabel berikut:

Tabel 5. Analisis variansi untuk peubah umur anak

Sumber variasi	Jumlah kuadrat	Derajat Bebas	Rataan Kuadrat	f hitung
Kelompok	347,370	1	347,370	140,709
Galat	40432,469	16378	2,469	
Total	40779,839	16379		

Dengan bantuan SPSS diperoleh $P_{value} = 0,000 < 0,05/7 = 0,007$

Kesimpulan : karena $P_{\text{value}} = 0,000 < 0,05/7 = 0,007$ dalam hal ini H_0 ditolak dan disimpulkan bahwa peubah status diare dan peubah umur anak adalah dependen.

Uji dependen independensi diare dengan jenis kelamin:

1. Perumusan hipotesis:
 $H_0 : P_{ij} = P_{i\bullet} \cdot P_{\bullet j}$ (Tidak terdapat hubungan antara diare dengan jenis kelamin)
 $H_1 : P_{ij} \neq P_{i\bullet} \cdot P_{\bullet j}$ (Terdapat hubungan antara diare dengan jenis kelamin)
2. Pengujian setiap pasangan kategori peubah merek hp dan peubah pembelian hp
3. Nilai statistik χ^2_{hitung} diperoleh sebagai berikut:

$$\begin{aligned}\chi^2_{\text{hitung}} &= \sum_{i=1}^b \sum_{j=1}^k \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \\ &= 24,364\end{aligned}$$

Berdasarkan tabel distribusi *chi-square* dengan $\alpha = 0,05$ dan $dk = 1$ diperoleh $\chi^2_{(0,05;1)} = 3,84$.

Karena $24,364 > 3,84$ maka H_0 tidak diterima artinya terdapat hubungan antara diare dengan jenis kelamin. Dengan bantuan GRETl diperoleh $P_{\text{value}} = 7,97451e - 007$.

Diperoleh hasil analisis menggunakan *software* SPSS dan GRETl dari masing-masing peubah bebas seperti yang ditampilkan pada Tabel 6.

Tabel 6. Hasil uji statistik pemilihan peubah penyekat simpul t_0

Peubah	Nama	Nilai Statistik Uji	P_{value}
X_1	Umur anak	$F = 140,709$	2,42644e-032*
X_2	Jenis kelamin	$\chi^2 = 24,364$	7,97451e-007
X_3	Sumber air minum	$\chi^2 = 38,600$	4,15065e-009
X_4	Fasilitas kakus	$\chi^2 = 40,837$	8,70067e-007
X_5	Daerah tempat tinggal	$\chi^2 = 16,347$	5,27396e-005
X_6	Tingkat pendidikan ibu	$F = 14,710$	1,25849e-010
X_7	Kuantil kekayaan	$F = 60,606$	7,38671e-015

Tabel 6 menunjukkan bahwa peubah yang mempunyai P_{value} terkecil adalah umur anak (X_1), karena $P_{\text{value}} = 2,42644e - 032 < \frac{\alpha}{7} = \frac{0,05}{7} = 0,007$ maka peubah X_1 dipilih sebagai peubah penyekat simpul t_0 .

2. Algoritma Pemilihan Titik Penyekat

Setelah melakukan pemilihan peubah penyekat selanjutnya menentukan titik penyekat pada pohon klasifikasi *QUEST*. Misalkan menentukan titik penyekat pada peubah penyekat umur anak (X_1) dalam penyekatan simpul t_0 pada pohon klasifikasi *QUEST*, sehingga memecah simpul t_0 menjadi dua simpul baru yaitu simpul t_1 dan simpul t_2 . Berdasarkan hasil perhitungan pada data diperoleh nilai $\bar{x}_1 < \bar{x}_2$ maka $d = \bar{x}_1$ sehingga titik penyekat simpul t_0 adalah $d = 3,5$. Hal ini menunjukkan bahwa peubah umur anak (X_1) yang memotong di nilai $d = 3,5$ yaitu umur anak 12-23 bulan. Dengan demikian, simpul t_0 disekat menjadi dua simpul, yaitu simpul t_1 dengan umur anak $\leq 12-23$ bulan dan simpul t_2 dengan umur anak $> 12-23$ bulan.

3. Algoritma Penghentian Pembentukan Pohon

Setelah dilakukan pemilihan peubah penyekat dan pemilihan titik penyekat secara *rekursif* selanjutnya penghentian pembentukan pohon. Proses penyekatan akan dihentikan apabila

banyaknya data pada simpul dalam paling sedikit 100, sedangkan banyaknya data pada simpul akhir paling sedikit 50, serta pertumbuhan pohon sudah mencapai kedalaman 5.

3.3 Ketepatan Klasifikasi Analisis *QUEST*

Hasil ketepatan klasifikasi menggunakan analisis *QUEST* diperoleh berdasarkan pohon dapat dilihat pada Tabel 7:

Tabel 7. Ketepatan prediksi klasifikasi *QUEST*

Observasi	Prediksi		
	Diare	Tidak Diare	Peresentase Ketepatan
Diare	885	1457	37,8%
Tidak Diare	0	14038	100,0%
Persentase Keseluruhan	5,4%	94,6%	91,1%

Persentase ketepatan prediksi diare (*sensitivity*) sebesar 37,8, sementara untuk ketepatan prediksi yang tidak diare (*specificity*) sebesar 100% dan untuk ketepatan keseluruhan prediksi yaitu 91,1%.

4 Kesimpulan

Berdasarkan hasil perhitungan dan pembahasan didapatkan beberapa kesimpulan sebagai berikut:

1. Prosedur klasifikasi berstruktur pohon biner dengan metode *QUEST* yang dilakukan untuk mengetahui faktor-faktor yang mempengaruhi penyakit diare yaitu:
 - a. Memilih penyekat dari tiap peubah bebas untuk memperoleh pemilah terbaik
 - b. Melakukan algoritma penentuan titik penyekat.
 - c. Penghentian pembentukan pohon dilakukan ketika
2. Faktor-faktor yang mempengaruhi penyakit diare pada balita dengan menggunakan metode *QUEST* adalah umur anak, fasilitas kakus, daerah tempat tinggal, pendidikan ibu, dan kuantil kekayaan. Ketepatan klasifikasi yang dihasilkan sebesar 91,1% atau risiko kesalahan sebesar 8,9% dengan nilai sensitivitas sebesar 37,8% dan nilai spesifisitas sebesar 100%.
3. Peubah yang paling berpengaruh pada balita diare adalah umur anak dengan kategori 12-23 bulan karena menjadi pemilah awal pada pohon klasifikasi *QUEST*.

Ucapan Terima Kasih

Penulis mengucapkan terima kasih kepada dosen pembimbing 1 Bapak Dr. Kartiko, M.Si., dan dosen pembimbing II Bapak Drs. Yudi Setyawan, MS., M.Sc., yang telah memberikan banyak waktu, tenaga, dan pikiran mulai dari awal sampai akhir laporan penelitian ini. Penulis juga mengucapkan terima kasih kepada kedua orang tua, saudara, serta teman-teman penulis yang selalu memberi dukungan, dorongan, serta doa dari awal sampai akhir penyusunan penelitian ini.

Daftar Pustaka

- Alwi, W., dan Tahir, N.A., 2014, Metode Klasifikasi Berstruktur Pohon dengan Algoritma Quest, Jurnal MSA, No. 2, Vol. 3, 2355-083X.
- Breiman L., Friedman JH., Olshen RA & Stone CJ., 1984, *Classification and Regression Trees*, Chapman and Hall, New York.
- Handayani, P., 2014 Gambaran Karakteristik Kejadian Diare Pada Balita Di Klinik Wijaya Kusuma Serpong Periode Januari – Desember 2013, STIKes Widya Dharma Husada.

- Irianto, J., Soesanto. S., Supraptini, Inswiasri, Irianti, S., dan Anwar, A., 1996, Faktor-Faktor yang Mempengaruhi Kejadian Diare pada Anak Balita (Analisis Lanjut Data SDKI 1994), *Buletin Penelitian Kesehatan*, Vol 24 (2 dan 3) 1996 : 77-96.
- Loh W-Y, Shih Y-S, 1997, Split Selection Methods for Classification Trees, *Statistica Sinica*7, 815-840.
- Septiani, T., 2016, Faktor-Faktor yang Mempengaruhi Pasien Penderita Stroke Di RSUD DR. Moewardi Menggunakan Pohon Klasifikasi Biner dengan Algoritma QUEST, *Skripsi*, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Sebelas Maret Surakarta.
- Suharyono, 2008, Diare Akut, PT Gramedia Pustaka Utama, Jakarta.
- Walpole, R. E., 1992, Pengantar Statistika Edisi Ke-4, PT Gramedia Pustaka Utama, Jakarta.
- World Health Organization, (2009), World Health Organization, *Identifying Priorities for Child Health Research to Achieve Millennium Development Goal 4: Consultation Proceedings.*, (March), 26–27. Retrieved from http://www.who.int/maternal_child_adolescent/documents/9789241598651/en/index.htm.
- . <http://www.depkes.go.id/resources/download/pusdatin/profil-kesehatan-indonesia/profil-kesehatan-Indonesia-2015.pdf>. *Profil Kesehatan Indonesia 2015*. (Diakses pada 23 Februari 2017).
- . 2013. *Survei Demografi dan Kesehatan Indonesia 2012*, Jakarta, BPS,BKKBN dan Departemen Kesehatan.