

**PEMBENTUKAN POHON KLASIFIKASI BINER DENGAN  
ALGORITMA QUEST (*QUICK, UNBIASED, AND EFFICIENT  
STATISTICAL TREE*) PADA DATA PASIEN LIVER**



---

---

**SKRIPSI**

---

---

**Disusun oleh  
MUHAMMAD ROSYID ABDURRAHMAN  
24010210120036**

**JURUSAN STATISTIKA  
FAKULTAS SAINS DAN MATEMATIKA  
UNIVERSITAS DIPONEGORO  
SEMARANG  
2014**

**PEMBENTUKAN POHON KLASIFIKASI BINER DENGAN  
ALGORITMA QUEST (*QUICK, UNBIASED, AND EFFICIENT  
STATISTICAL TREE*) PADA DATA PASIEN LIVER**

**MUHAMMAD ROSYID ABDURRAHMAN  
24010210120036**

Skripsi

Diajukan Sebagai Syarat untuk Memperoleh Gelar Sarjana Sains  
pada Jurusan Statistika

**JURUSAN STATISTIKA  
FAKULTAS SAINS DAN MATEMATIKA  
UNIVERSITAS DIPONEGORO  
SEMARANG**

**2014**

## HALAMAN PENGESAHAN I

Judul : Pembentukan Pohon Klasifikasi Biner dengan Algoritma QUEST

(*Quick, Unbiased, and Efficient Statistical Tree*) pada Data Pasien Liver

Nama : Muhammad Rosyid Abdurrahman

NIM : 24010210120036

Telah diajukan pada sidang Tugas Akhir tanggal 17 September 2014 dan dinyatakan lulus pada tanggal 25 September 2014.

Semarang, 26 September 2014

Mengetahui,

Ketua Jurusan Statistika

FSM UNDIP,  
  
Dia. Daw Ispryanti, M.Si  
NIP. 195709141986032001

Ketua Panitia Penguji

Ujian Tugas Akhir,



Prof. Drs. Mustafid, M.Eng, PhD

NIP. 195505281980031002

## HALAMAN PENGESAHAN II

Judul : Pembentukan Pohon Klasifikasi Biner dengan Algoritma QUEST

(*Quick, Unbiased, and Efficient Statistical Tree*) pada Data Pasien Liver

Nama : Muhammad Rosyid Abdurrahman

NIM : 24010210120036

Telah diajukan pada sidang Tugas Akhir tanggal 17 September 2014 dan dinyatakan lulus pada tanggal 25 September 2014.

Semarang, 26 September 2014

Pembimbing I



Dra. Dwi Ispriyanti, M.Si

NIP. 195709141986032001

Pembimbing II



Alan Prahutama, M.Si

## KATA PENGANTAR

Puji syukur kehadiran Allah SWT yang telah memberikan rahmat, hidayah, dan karunia-Nya sehingga penulis dapat menyelesaikan penulisan skripsi dengan judul “**Pembentukan Pohon Klasifikasi Biner dengan Algoritma QUEST (*Quick, Unbiased, and Efficient Statistical Tree*) pada Data Pasien Liver**”.

Penulis menyadari skripsi ini tidak akan dapat diselesaikan tanpa bantuan dari berbagai pihak. Oleh karena itu, penulis ingin menyampaikan terima kasih kepada:

1. Ibu Dra. Dwi Ispriyanti, M.Si selaku Ketua Jurusan Statistika Fakultas Sains dan Matematika Universitas Diponegoro dan sekaligus dosen pembimbing I.
2. Bapak Alan Prahutama, M.Si selaku dosen pembimbing II yang telah berkenan meluangkan waktu dalam memberikan masukan, arahan, dan bimbingan kepada penulis.
3. Seluruh Bapak dan Ibu Dosen Jurusan Statistika FSM Universitas Diponegoro yang telah memberikan ilmu yang sangat berguna.
4. Semua pihak yang telah membantu dalam penulisan skripsi ini yang tidak dapat disebutkan satu per satu.

Penulis menyadari bahwa skripsi ini masih jauh dari kesempurnaan. Oleh karena itu kritik dan saran yang sifatnya membangun sangat penulis harapkan demi kesempurnaan penulisan selanjutnya.

Semarang, September 2014

**Penulis**

## ABSTRAK

Di era modern ini semakin banyak makanan cepat saji yang terkadang mempunyai kandungan zat kimia dan semakin banyaknya kendaraan bermotor yang tidak terkendali peredarannya menyebabkan polusi udara sehingga dapat mempengaruhi kesehatan organ hati manusia. Untuk membantu dalam menganalisis ada atau tidaknya gangguan hati pada manusia digunakan algoritma QUEST (*Quick, Unbiased, and Efficient Statistical Tree*) untuk mengklasifikasikan ciri-ciri pasien liver berdasarkan uji fungsi hati yang dilakukan di laboratorium klinik. QUEST membentuk aturan untuk memprediksi kelas dari sebuah objek dari nilai-nilai variabel prediktor. Pohon dibentuk melalui penyekatan (mempartisi) data secara berulang, di mana kelas dan nilai-nilai variabel prediktor setiap amatan pada data contoh sudah diketahui. Masing-masing sekatan (*split*) data dinyatakan sebagai simpul (*node*) dalam pohon yang terbentuk. QUEST merupakan salah satu metode pohon klasifikasi biner. Hasil dari pohon klasifikasi yang terbentuk, variabel penting dalam mengklasifikasikan seseorang terkena penyakit liver atau tidak, yaitu variabel *Direct Bilirubin*, *Fosfatase Alkali*, *Serum Glutamic Oxaloacetic Transaminase* (SGOT), dan Umur pasien. Akurasi dari algoritma QUEST mengklasifikasi data pasien liver dalam skripsi ini sebesar 73,4 %.

**Kata kunci** : pohon klasifikasi biner, algoritma QUEST, data pasien liver.

## ABSTRACT

In this modern era of fast food commonly found that sometimes have chemical substances and the increasing number of motor vehicles that cause the uncontrolled circulation of air pollution that can affect the health of the human liver. To assist in analyzing the presence of liver disorders in humans can be used QUEST (Quick, Unbiased, and Efficient Statistical Tree) algorithm to classify the characteristics of the patient's liver by liver function tests performed in clinical laboratories. QUEST construct rules to predict the class of an object from the values of predictor variables. The tree is constructed by partitioning the data by recursively, where class and the values of the predictor variables of each observation in the data sample is known. Each partition is represented by a node in the tree. QUEST is one of the binary classification tree method. The results of the classification tree is formed, an important variable in classifying a person affected by liver disease or not, that is the variable Direct Bilirubin, Alkaline Phosphatase, Serum Glutamic Oxaloacetic Transaminase (SGOT), and age of the patient. Accuracy of the QUEST algorithm classifying liver patient data by 73,4 %.

**Keywords:** binary classification trees, QUEST algorithm, liver patient data.

## DAFTAR ISI

	<b>Halaman</b>
HALAMAN JUDUL.....	i
HALAMAN PENGESAHAN I .....	ii
HALAMAN PENGESAHAN II.....	iii
KATA PENGANTAR .....	iv
ABSTRAK .....	v
ABSTRACT .....	vi
DAFTAR ISI.....	vii
DAFTAR TABEL.....	x
DAFTAR GAMBAR .....	xi
DAFTAR LAMPIRAN.....	xii
<b>BAB I     PENDAHULUAN</b>	
1.1. Latar Belakang.....	1
1.2. Permasalahan .....	3
1.3. Pembatasan Masalah.....	3
1.4. Tujuan Penulisan .....	3
<b>BAB II    TINJAUAN PUSTAKA</b>	
2.1. Tinjauan Umum Hati .....	4
2.2. Data Mining .....	7
2.2.1. Pohon Keputusan ( <i>Decision Tree</i> ).....	9
2.2.1.1. Algoritma QUEST.....	12
2.2.1.1.1. Algoritma Pemilihan Variabel Penyekat.....	13



	2.2.1.1.2. Algoritma Menentukan Titik Sekat.....	15
	2.2.1.1.3. Proses Pemberhentian Penyekatan Simpul .....	22
	2.3. Ketepatan Pohon Klasifikasi.....	23
BAB III	METODOLOGI PENELITIAN	
	3.1. Sumber Data .....	25
	3.2. Variabel Data .....	25
	3.3. Metode Analisis Data .....	26
BAB IV	HASIL DAN PEMBAHASAN	
	4.1. Proses Pembentukan Pohon Klasifikasi Biner.....	28
	4.1.1. Pemilihan Variabel Penyekat untuk Simpul 0 ( $t_0$ ) .....	29
	4.1.1.1. Perhitungan Manual untuk Variabel Prediktor Kontinu... 29	
	4.1.1.2. Perhitungan Manual untuk Variabel Prediktor Kategorik 30	
	4.1.2. Pemilihan Titik Sekat untuk Simpul 0 ( $t_0$ ) .....	31
	4.1.3. Pemilihan Variabel Penyekat untuk Simpul 1 ( $t_1$ ) .....	33
	4.1.4. Pemilihan Titik Sekat untuk Simpul 1 ( $t_1$ ) .....	33
	4.1.5. Pemilihan Variabel Penyekat untuk Simpul 3 ( $t_3$ ) .....	35
	4.1.6. Pemilihan Titik Sekat untuk Simpul 3 ( $t_3$ ) .....	36
	4.1.7. Pemilihan Variabel Penyekat untuk Simpul 7 ( $t_7$ ) .....	37
	4.1.8. Pemilihan Titik Sekat untuk Simpul 7 ( $t_7$ ) .....	38
	4.1.9. Pemilihan Variabel Penyekat untuk Simpul 13 ( $t_{13}$ ).....	40
	4.1.10. Pemilihan Variabel Penyekat untuk Simpul 8 ( $t_8$ ) .....	41
	4.2. Interpretasi Pohon Klasifikasi yang Terbentuk .....	44
	4.2.1. Aturan Klasifikasi yang Terbentuk.....	44
	4.2.2. Ukuran Ketepatan Prediksi.....	47

BAB V	PENUTUP	
5.1.	Kesimpulan .....	48
5.2.	Saran .....	49
DAFTAR PUSTAKA	.....	50

## DAFTAR TABEL

	<b>Halaman</b>
<b>Tabel 1</b> Matriks Konfusi.....	23
<b>Tabel 2</b> Variabel Data.....	25
<b>Tabel 3</b> Hasil Uji Statistika Pemilihan Variabel Penyekat Simpul 0 .....	29
<b>Tabel 4</b> Tabel Tabulasi Silang untuk Variabel Jenis Kelamin .....	30
<b>Tabel 5</b> Hasil Uji Statistika Pemilihan Variabel Penyekat Simpul 1 .....	33
<b>Tabel 6</b> Hasil Uji Statistika Pemilihan Variabel Penyekat Simpul 3 .....	35
<b>Tabel 7</b> Hasil Uji Statistika Levene F Pemilihan Variabel Penyekat Simpul 3 ...	35
<b>Tabel 8</b> Hasil Uji Statistika Pemilihan Variabel Penyekat Simpul 7 .....	37
<b>Tabel 9</b> Hasil Uji Statistika Levene F Pemilihan Variabel Penyekat Simpul 7 ...	38
<b>Tabel 10</b> Hasil Uji Statistika Pemilihan Variabel Penyekat Simpul 13 .....	40
<b>Tabel 11</b> Hasil Uji Statistika Levene F Pemilihan Variabel Penyekat Simpul 13	40
<b>Tabel 12</b> Hasil Uji Statistika Pemilihan Variabel Penyekat Simpul 8 .....	41
<b>Tabel 13</b> Hasil Uji Statistika Levene F Pemilihan Variabel Penyekat Simpul 8 .	42
<b>Tabel 14</b> Matriks Konfusi Hasil Klasifikasi.....	47

## DAFTAR GAMBAR

	<b>Halaman</b>
<b>Gambar 1</b> Pohon Klasifikasi Biner .....	11
<b>Gambar 2</b> Diagram Alir Pengolahan Data.....	27
<b>Gambar 3</b> Pemecahan Simpul 0 ( $t_0$ ) .....	32
<b>Gambar 4</b> Pemecahan Simpul 1 ( $t_1$ ) .....	34
<b>Gambar 5</b> Pemecahan Simpul 3 ( $t_3$ ) .....	37
<b>Gambar 6</b> Pemecahan Simpul 7 ( $t_7$ ) .....	39
<b>Gambar 7</b> Pemecahan Simpul 6 ( $t_6$ ) .....	42
<b>Gambar 8</b> Pohon Klasifikasi Biner yang Terbentuk.....	43

## DAFTAR LAMPIRAN

	<b>Halaman</b>
<b>Lampiran 1</b> Data Pasien Liver.....	52
<b>Lampiran 2</b> Pohon Klasifikasi yang Terbentuk.....	53
<b>Lampiran 3</b> Hasil Klasifikasi Data Pasien Liver .....	56

# BAB I

## PENDAHULUAN

### 1.1. Latar Belakang

Hati merupakan salah satu organ terbesar dalam tubuh manusia yang mempunyai banyak fungsi bagi tubuh. Organ ini mempunyai fungsi yang kompleks, sehingga mudah terpengaruh gangguan penyakit. Oleh karena memiliki fungsi yang kompleks, tidak mudah dalam mendiagnosis penyakit liver (gangguan fungsi hati).

Dalam mendiagnosis ada atau tidak penyakit liver dapat digunakan acuan dari hasil tes fungsi hati yang dilaksanakan di laboratorium. Tes tersebut antara lain yaitu *transaminase serum*, fosfatase alkali, total bilirubin, bilirubin terkonjugasi, total protein, albumin, serta rasio albumin dan globulin. Dari hasil tes tersebut dapat dilihat hasil tes yang signifikan sebagai ciri adanya gangguan fungsi hati dengan menggunakan algoritma pohon klasifikasi karena dapat memperoleh informasi mengenai data klasifikasi pasien penyakit liver.

Menurut Rokach dan Maimon (2008) dan Maroco *et al* (2011), algoritma pohon klasifikasi merupakan pendekatan nonparametrik. Metode tersebut adalah salah satu metode klasifikasi yang tidak bergantung pada asumsi tertentu dan mampu mengeksplorasi struktur data yang kompleks dengan variabel yang banyak. Struktur data dapat dilihat secara visual dan algoritma pohon klasifikasi juga memberikan kemudahan dalam menginterpretasikan hasil.

Menurut Kim dan Loh (2001), algoritma pohon klasifikasi dibagi menjadi dua kelompok. Kelompok pertama yaitu pohon klasifikasi yang menghasilkan dua

simpul (biner) setiap penyekatan (*split*). Pohon klasifikasi biner terdiri dari dua, yaitu *Classification and Regression Tree* (CART) dan *Quick, Unbiased, and Efficient Statistical Tree* (QUEST). Kelompok kedua disebut pohon klasifikasi non-biner atau *multiway splits*, karena setiap penyekatan menghasilkan lebih dari dua simpul. Pohon klasifikasi non biner yaitu *Fast and Automatic Classification Tree* (FACT), C4.5, *Chi-Squared Automatic Interaction Detection* (CHAID), FIRM, dan CRUISE.

Algoritma QUEST merupakan algoritma pohon klasifikasi yang menghasilkan variabel tak bias dan memiliki dua simpul setiap penyekatan. QUEST dapat diterapkan pada data dengan variabel respon dua kategori berupa data nominal dan variabel prediktor dengan variabel berbentuk kategorik maupun kontinu. Pembentukan pohon klasifikasi dengan QUEST melalui dua langkah, yakni pemilihan variabel penyekat dan menentukan titik sekat. Pemilihan variabel penyekat digunakan uji ANOVA F, Levene F, dan chi-kuadrat Pearson dan dipilih variabel yang signifikan dengan *p-value* terkecil. Pemilihan variabel penyekat digunakan untuk menentukan titik sekat, yakni suatu nilai yang dapat mempartisi atau membagi data ke dalam dua simpul yang berbeda.

Penelitian dengan menggunakan algoritma pohon klasifikasi biner pernah dibahas oleh Krisan Aprian Widagdo pada tahun 2010 tentang penerapan pohon klasifikasi biner menggunakan algoritma CART pada kasus penyakit diabetes suku Pima Indian. Pada penelitian ini penulis membahas tentang pembentukan pohon klasifikasi biner dengan algoritma QUEST untuk mengetahui ciri-ciri pasien penyakit liver melalui hasil tes fungsi hati.

## 1.2. Permasalahan

Berdasarkan uraian pada latar belakang penelitian, permasalahan yang diangkat dalam skripsi ini adalah :

1. Bagaimana penerapan algoritma QUEST dalam pembentukan pohon klasifikasi biner untuk mengetahui ciri-ciri pasien penyakit liver dengan menggunakan hasil tes fungsi hati?
2. Bagaimana hasil klasifikasi yang dihasilkan dari pohon klasifikasi biner menggunakan algoritma QUEST, sehingga dapat diketahui akurasi pohon klasifikasi dalam mengklasifikasi?

## 1.3. Pembatasan Masalah

Pada penelitian ini, dilakukan pembatasan masalah yaitu dalam mengklasifikasi untuk menentukan ciri-ciri pasien penyakit liver digunakan variabel jenis kelamin pasien, umur pasien, total bilirubin, bilirubin terkonjugasi (*Direct Bilirubin*), *Fosfatase Alkali*, *Serum Glutamic Pyruvic Transaminase* (SGPT), *Serum Glutamic Oxaloacetic Transaminase* (SGOT), total protein, albumin, serta rasio albumin globulin.

## 1.4. Tujuan Penelitian

Tujuan penulisan tugas akhir ini adalah :

1. Membentuk pohon klasifikasi biner untuk mengetahui ciri-ciri pasien penyakit liver dengan menggunakan hasil tes fungsi hati.
2. Mengetahui tingkat akurasi pohon klasifikasi dalam mengklasifikasikan data.