

**PENERAPAN METODE *EXHAUSTIVE CHI-SQUARE*
AUTOMATIC INTERACTION DETECTION PADA
KLASIFIKASI PENDERITA DIABETES DAN NON
DIABETES**

SKRIPSI



NURHIDAYATULLAH

H051171019

**PROGRAM STUDI STATISTIKA DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM**

UNIVERSITAS HASANUDDIN

MAKASSAR

NOVEMBER 2023

**PENERAPAN METODE *EXHAUSTIVE CHI-SQUARE*
AUTOMATIC INTERACTION DETECTION PADA
KLASIFIKASI PENDERITA DIABETES DAN NON
DIABETES**

SKRIPSI

Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana Sains pada
Program Studi Statistika Departemen Statistika Fakultas Matematika dan Ilmu
Pengetahuan Alam Universitas Hasanuddin

NURHIDAYATULLAH

H051171019

**PROGRAM STUDI STATISTIKA DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS HASANUDDIN
MAKASSAR
NOVEMBER 2023**

LEMBAR PERNYATAAN KEOTENTIKAN

Saya yang bertanda tangan di bawah ini menyatakan dengan sungguh-sungguh bahwa skripsi yang saya buat dengan judul:

Penerapan Metode *Exhaustive Chi-Square Automatic Interaction Detection* pada Klasifikasi Penderita Diabetes dan Non Diabetes

adalah benar hasil karya saya sendiri, bukan hasil plagiat dan belum pernah dipublikasikan dalam bentuk apapun

Makassar, 29 November 2023



Nurhidayatullah

NIM H051171019

PENERAPAN METODE *EXHAUSTIVE CHI-SQUARE*
AUTOMATIC INTERACTION DETECTION PADA
KLASIFIKASI PENDERITA DIABETES DAN NON DIABETES

Disetujui Oleh:

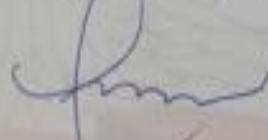
Pembimbing Utama



Siti Sahriman, S.Si., M.Si.

NIP. 19881018 201504 2 002

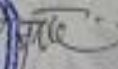
Pembimbing Pendamping



Dr. Nirwan, M.Si.

NIP. 19630306 198702 1 002

Ketua Program Studi



Nurhayati, S.Si., M.Si.

NIP. 19770808 200501 2 002

Pada 29 November 2023

HALAMAN PENGESAHAN

Skripsi ini diajukan oleh:

Nama : Nurhidayatullah
NIM : H051171019
Program Studi : Statistika
Judul Skripsi : Penerapan Metode *Exhaustive Chi-Square Automatic Interaction Detection* pada Klasifikasi Penderita Diabetes dan Non Diabetes

Telah berhasil dipertahankan dihadapan Dewan Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Sains pada Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin.

DEWAN PENGUJI

1. Ketua : Sitti Sahriman, S.Si., M.Si. (.....)
2. Sekretaris : Dr. Nirwan, M.Si. (.....)
3. Anggota : Dr. Erna Tri Herdiani, S.Si., M.Si. (.....)
4. Anggota : Sri Astuti Thamrin, S.Si., M.Stat., Ph.D (.....)

Ditetapkan di : Makassar

Tanggal : 29 November 2023

KATA PENGANTAR

AlhamdulillahRobbil'alamin, Segala Puji senantiasa disenandungkan untuk menyampaikan rasa syukur penulis kepada **Allah Subhanahu Wata'ala** atas segala limpahan rahmat, nikmat, dan hidayah yang diberikan kepada penulis sehingga dapat menyelesaikan penulisan skripsi dengan judul **“Penerapan Metode Exhaustive Chi-Square Automatic Interaction Detection pada Klasifikasi Penderita Diabetes dan Non Diabetes”** sebagai salah satu syarat untuk mendapat gelar Sarjana Sains pada Program Studi Statistika Departemen Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin Makassar.

Penulis menyadari dalam penulisan skripsi ini tidak lepas dari hambatan dan masalah namun dapat terselesaikan berkat dukungan dan bantuan berbagai pihak. Oleh karena itu, penulis menyampaikan rasa terima kasih yang setulus-tulusnya serta penghargaan yang setinggi-tingginya untuk orang tua penulis, Ayahanda **Nurdin Rahimahullah** dan Ibunda tercinta **Hariati**, yang telah membesarkan dan mendidik dengan penuh kesabaran, memberikan cinta dan limpahan kasih sayang, dukungan dan doa yang tulus tanpa henti kepada penulis. Rasa terima kasih juga kepada kakak tersayang **Nurhana** dan **Sudarmin**, adik tersayang **Rahmawati** dan keponakan **Ahmad Rahul Taufik**. Tak lupa pula terima kasih untuk **Keluarga Besar** atas doa, dukungan, semangat, dan bantuannya kepada penulis.

Penghargaan yang tulus dan ucapan terima kasih dengan penuh keikhlasan juga penulis ucapkan kepada:

1. **Bapak Prof. Dr. Ir. Jamaluddin Jompa, M.Sc.**, selaku Rektor Universitas Hasanuddin beserta seluruh jajarannya.
2. **Bapak Dr. Eng. Amiruddin, M.Si.** selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin beserta seluruh jajarannya.
3. **Ibu Dr. Anna Islamiyati, S.Si., M.Si.** selaku Ketua Departemen Statistika, segenap Dosen Pengajar dan Staf Departemen Statistika yang telah membekali ilmu dan kemudahan-kemudahan kepada Penulis dalam berbagai hal selama menjadi mahasiswa di Departemen Statistika.

4. **Ibu Dr. Erna Tri Herdiani, S.Si., M.Si.** selaku Penasehat Akademik atas saran, nasehat, motivasi, dan dukungan yang telah diberikan kepada penulis selama menjadi mahasiswa dan meluangkan waktu untuk membimbing dan memberikan masukan dalam penulisan skripsi ini.
5. **Ibu Sitti Sahrman, S.Si., M.Si.** selaku Pembimbing Utama dan **Bapak Dr. Nirwan, M.Si.** selaku Pembimbing Pertama yang telah ikhlas meluangkan waktu dan pemikirannya untuk memberikan arahan, pengetahuan, motivasi dan bimbingan ditengah kesibukan beliau serta menjadi tempat berkeluh kesah untuk penulis.
6. **Ibu Dr. Erna Tri Herdiani, S.Si., M.Si.** dan **Ibu Sri Astuti Thamrin, S.Si., M.Stat., Ph.D.** selaku Tim Penguji atas saran dan kritikan yang membangun dalam penyempurnaan penyusunan skripsi ini serta waktu yang telah diberikan kepada Penulis.
7. Sahabat seperjuangan dalam penyusunan skripsi, **Zakiah Fitri** dan **Sriulan S.S. Abdullah** yang selalu setia membersamai penulis, memberikan dukungan satu sama lain dan menjadi tempat bertukar cerita. Terima kasih juga kepada **Kak Muhammad Fadil** untuk segala bantuan dan motivasi yang tidak ada putusnya.
8. Sahabat tercinta penulis, **Fakhriyyah Dj Junus, Nurul Wahyuni, Nurul Annisa, Riska Rasyid, Nur Aprilia Dzulhijjah, Sakinah Oktoni, Miftahul Jannah, Munadiah Apriliani, Risnawati Azali** dan **Fitri** yang telah menjadi sahabat terbaik sejak awal perkuliahan hingga saat ini.
9. Teman-teman **Statistika 2017**, terima kasih atas kebersamaan, suka, dan duka selama menjalani pendidikan di Departemen Statistika. Penulis senang mengenal kalian semua. Terkhusus untuk **Aqilah, Haura, Ata, Zulfa, Ukhty, Ica, Mirna, Putri, Wiwi, Iklil, Aii** dan **Rahman** terima kasih atas segala bantuan dan motivasinya untuk penulis dalam menyelesaikan skripsi ini.
10. Keluarga besar **DISKRIT 2017**, terima kasih telah memberikan pelajaran yang berharga dan arti kebersamaan selama ini kepada penulis. Pengalaman yang berharga telah penulis dapatkan dari teman-teman selama berproses.
11. **Keluarga Mahasiswa FMIPA Unhas** terkhusus anggota keluarga **Himatika FMIPA Unhas** dan **Himastat FMIPA Unhas**, terima kasih atas ilmu yang

mungkin tidak bisa didapatkan di proses perkuliahan dan telah menjadi keluarga selama penulis kuliah di Universitas Hasanuddin.

12. Semua pihak yang telah banyak berpartisipasi, baik secara langsung maupun tidak langsung yang tak sempat penulis sebutkan satu per satu. Terima kasih untuk segala bantuan dan dukungannya.

Penulis menyadari bahwa masih banyak kekurangan dalam skripsi ini. Untuk itu dengan segala kerendahan hati penulis memohon maaf. Akhir kata, semoga skripsi ini dapat memberikan manfaat bagi pembaca.

Wassalamu'alaikum Warahmatullahi Wabarakatuh

Makassar, 29 November 2023



Nurhidayatullah

PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS

Sebagai civitas akademik Universitas Hasanuddin, saya yang bertanda tangan di bawah ini:

Nama : Nurhidayatullah
NIM : H051171019
Program Studi : Statistika
Departemen : Statistika
Fakultas : Matematika dan Ilmu Pengetahuan Alam
Jenis Karya : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Hasanuddin **Hak Bebas Royalti Non-eksklusif (*Non-exclusive Royalty- Free Right*)** atas tugas akhir saya yang berjudul:

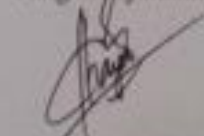
**“Penerapan Metode Exhaustive Chi-Square Automatic Interaction Detection
pada Klasifikasi Penderita Diabetes dan Non Diabetes”**

Beserta perangkat yang ada (jika diperlukan). Terkait dengan hal di atas, maka pihak universitas berhak menyimpan, mengalih-media/format-kan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di Makassar pada tanggal 29 November 2023.

Yang menyatakan



(Nurhidayatullah)

ABSTRAK

Klasifikasi merupakan suatu proses pengelompokan suatu objek ke dalam kategori tertentu. Salah satu metode klasifikasi yang dapat digunakan pada data kategorik yaitu metode *Exhaustive Chi-Square Automatic Interaction Detection* (CHAID). Metode *Exhaustive* CHAID adalah metode klasifikasi untuk data kategorik dengan membentuk pohon klasifikasi. Metode *Exhaustive* CHAID terdiri dari tiga tahap yaitu tahap penggabungan, pemisahan, dan penghentian. Pohon klasifikasi yang diperoleh menginterpretasikan variabel-variabel prediktor yang berpengaruh signifikan terhadap variabel respon berdasarkan uji *chi-square*. Tujuan dari penelitian ini adalah untuk mendapatkan hasil klasifikasi penderita diabetes dan non diabetes menggunakan metode *Exhaustive* CHAID. Variabel respon yang digunakan yaitu kadar gula darah masyarakat penderita diabetes dan non diabetes di Kabupaten Muna Barat. Variabel prediktor terdiri dari tekanan darah sistolik, tekanan darah diastolik, lama waktu tidur, gaya bekerja, tingkat pengetahuan tentang diabetes, ukuran lingkar perut, riwayat keturunan diabetes, usia, kebiasaan berolahraga, dan indeks massa tubuh. Hasil klasifikasi metode *Exhaustive* CHAID yang diperoleh menunjukkan bahwa faktor-faktor yang berpengaruh signifikan pada taraf 5% yaitu riwayat keturunan diabetes, ukuran lingkar perut, tingkat pengetahuan tentang diabetes, dan tekanan darah diastolik. Selain itu, diperoleh nilai akurasi dari pohon klasifikasi *Exhaustive* CHAID yang cukup baik yakni sebesar 86% berdasarkan matriks konfusi.

Kata Kunci: Klasifikasi, Pohon Klasifikasi, Uji *Chi-Square*, *Exhaustive Chi-Square Automatic Interaction Detection*, Diabetes

ABSTRACT

Classification is a process of grouping an object into a certain category. One of classification method that can be used on categorical data is the Exhaustive Chi-Square Automatic Interaction Detection (CHAID) method. The Exhaustive CHAID method is a classification method for categorical data by forming a classification tree. The Exhaustive CHAID method consists of three stages, namely the merging, separation and termination stages. The classification tree obtained interprets the predictor variables that have a significant effect on the response variable based on the chi-square test. The purpose of this study was to obtain classification results for diabetics and non-diabetics using the Exhaustive CHAID method. The response variable used is the blood sugar levels of people with diabetics and non-diabetics in West Muna Regency. Predictor variables consist of systolic blood pressure, diastolic blood pressure, length of sleep, working style, level of knowledge about diabetes, abdominal circumference, hereditary history of diabetes, age, exercise habits, and body mass index. The results of the Exhaustive CHAID classification method obtained show that the factors that have a significant influence at the 5% level are a hereditary history of diabetes, abdominal circumference, level of knowledge about diabetes, and diastolic blood pressure. Apart from that, the accuracy value obtained from the Exhaustive CHAID classification tree was quite good, namely 86% based on the confusion matrix.

Keywords: Classification, Classification Tree, Chi-Square Test, Exhaustive Chi-Square Automatic Interaction Detection, Diabetes

DAFTAR ISI

HALAMAN JUDUL.....	i
LEMBAR PERNYATAAN KEOTENTIKAN.....	ii
LEMBAR PERSETUJUAN PEMBIMBING	iii
HALAMAN PENGESAHAN.....	iv
KATA PENGANTAR	v
PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS	viii
ABSTRAK	ix
ABSTRACT.....	x
DAFTAR ISI.....	xi
DAFTAR GAMBAR	xiii
DAFTAR TABEL	xiv
DAFTAR LAMPIRAN.....	xv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah.....	3
1.4 Tujuan Penelitian.....	3
1.5 Manfaat Penelitian.....	4
BAB II TINJAUAN PUSTAKA.....	5
2.1 Metode <i>Exhaustive</i> CHAID.....	5
2.2 Variabel pada Metode <i>Exhaustive</i> CHAID	5
2.3 Uji <i>Chi-Square</i>	6
2.4 Koreksi <i>Bonferroni</i>	8
2.5 Algoritma Metode <i>Exhaustive</i> CHAID	9
2.5.1 Tahap Penggabungan	9
2.5.2 Tahap Pemisahan.....	9
2.5.3 Tahap Penghentian	10
2.6 Matriks Konfusi.....	10
2.7 Diabetes Melitus.....	11
BAB III METODOLOGI PENELITIAN	13

3.1 Sumber Data	13
3.2 Variabel Penelitian	13
3.3 Metode Analisis Data	14
BAB IV HASIL DAN PEMBAHASAN	16
4.1 Karakteristik Data.....	16
4.1.1 Tekanan Darah Sistolik	16
4.1.2 Tekanan Darah Diastolik.....	17
4.1.3 Lama Waktu Tidur	18
4.1.4 Gaya Bekerja	18
4.1.5 Tingkat Pengetahuan tentang Diabetes	19
4.1.6 Ukuran Lingkar Perut.....	20
4.1.7 Riwayat Keturunan Diabetes.....	20
4.1.8 Usia.....	21
4.1.9 Kebiasaan Berolahraga.....	22
4.1.10 Indeks Massa Tubuh.....	22
4.2 Klasifikasi Penderita Diabetes dan Non Diabetes menggunakan Metode <i>Exhaustive CHAID</i>	23
4.2.1 Tahap Penggabungan	23
4.2.2 Tahap Pemisahan.....	28
4.2.3 Tahap Penghentian	33
4.2.4 Keakuratan Klasifikasi Metode <i>Exhaustive CHAID</i>	38
BAB V KESIMPULAN DAN PEMBAHASAN	40
5.1 Kesimpulan.....	40
5.2 Saran	40
DAFTAR PUSTAKA	41
LAMPIRAN.....	44

DAFTAR GAMBAR

Gambar 4.1 Frekuensi Tekanan Darah Sistolik Berdasarkan Kadar Gula Darah	17
Gambar 4.2 Frekuensi Tekanan Darah Diastolik Berdasarkan Kadar Gula Darah.....	17
Gambar 4.3 Frekuensi Lama Waktu Tidur Berdasarkan Kadar Gula Darah.....	18
Gambar 4.4 Frekuensi Gaya Bekerja Berdasarkan Kadar Gula Darah	19
Gambar 4.5 Frekuensi Tingkat Pengetahuan tentang Diabetes Berdasarkan Kadar Gula Darah	19
Gambar 4.6 Frekuensi Ukuran Lingkar Perut Berdasarkan Kadar Gula Darah ...	20
Gambar 4.7 Frekuensi Riwayat Keturunan Diabetes Berdasarkan Kadar Gula Darah.....	21
Gambar 4.8 Frekuensi Usia Berdasarkan Kadar Gula Darah.....	21
Gambar 4.9 Frekuensi Kebiasaan Berolahraga Berdasarkan Kadar Gula Darah..	22
Gambar 4.10 Frekuensi Indeks Massa Tubuh Tekanan Darah Sistolik Berdasarkan Kadar Gula Darah.....	23
Gambar 4.11 Pohon Klasifikasi <i>Exhaustive</i> CHAID.....	34

DAFTAR TABEL

Tabel 2.1 Struktur Data Uji <i>Chi-Square</i>	6
Tabel 2.2 Matriks Konfusi.....	11
Tabel 3.1 Variabel Penelitian	13
Tabel 4.1 Frekuensi Kadar Gula Darah.....	16
Tabel 4.2 Tabel Kontingensi antara Kadar Gula Darah dengan Kategori Satu Tekanan Darah Sistolik	24
Tabel 4.3 Frekuensi Harapan antara Kadar Gula Darah dengan Kategori Satu Tekanan Darah Sistolik	24
Tabel 4.4 Hasil Uji <i>Chi-Square</i> antara Kadar Gula Darah dengan Kategori Tekanan Darah Sistolik	25
Tabel 4.5 Hasil Uji <i>Chi-Square</i> dan Penggabungan Kategori.....	26
Tabel 4.6 Nilai Koreksi <i>Bonferroni</i> Variabel Prediktor	27
Tabel 4.7 Nilai <i>p-value</i> terkoreksi <i>Bonferroni</i> untuk Variabel Prediktor.....	27
Tabel 4.8 Hasil Uji <i>Chi-Square</i> antara Kadar Gula Darah dengan Variabel Prediktor	28
Tabel 4.9 Hasil Uji <i>Chi-Square</i> antara Kadar Gula Darah dengan Variabel Prediktor untuk Cabang <i>Node 1</i>	30
Tabel 4.10 Hasil Uji <i>Chi-Square</i> antara Kadar Gula Darah dengan Variabel Prediktor untuk Cabang <i>Node 2</i>	31
Tabel 4.11 Hasil Uji <i>Chi-Square</i> antara Kadar Gula Darah dengan Variabel Prediktor untuk Cabang <i>Node 5</i>	32
Tabel 4.12 Segmen Kadar Gula Darah.....	37
Tabel 4.13 Presentase Segmen Kadar Gula Darah	37
Tabel 4.14 Matriks Konfusi Klasifikasi <i>Exhaustive CHAID</i>	38

DAFTAR LAMPIRAN

Lampiran 1 Data Penderita Diabetes dan Non Diabetes di Kabupaten Muna Barat.....	45
Lampiran 2 Hasil Uji <i>Chi-square</i> antara Kadar Gula Darah dengan masing-masing Kategori Variabel Prediktor	47
Lampiran 3 Hasil Uji <i>Chi-square</i> antara Kadar Gula Darah dengan Variabel Prediktor untuk Cabang <i>Node 3</i> dan <i>Node 4</i>	49
Lampiran 4 Hasil Uji <i>Chi-square</i> antara Kadar Gula Darah dengan Variabel Prediktor untuk Cabang <i>Node 6</i>	50
Lampiran 5 Hasil Prediksi Kadar Gula Darah Metode <i>Exhaustive CHAID</i>	51

BAB I

PENDAHULUAN

1.1 Latar Belakang

Diabetes melitus merupakan salah satu masalah kesehatan masyarakat yang menjadi ancaman serius yang dihadapi dunia. Angka kejadian penyakit diabetes melitus meningkat secara drastis di negara berkembang, termasuk di Indonesia (Dewi, 2018). Diabetes melitus adalah suatu kondisi kronis yang terjadi ketika tubuh tidak dapat menghasilkan hormon *polipeptida* yang cukup untuk mengatur metabolisme (Azis dkk., 2020). Diabetes melitus disebut juga *the silent killer* sebab penyakit ini dapat menyerang beberapa organ tubuh yang mengakibatkan berbagai macam keluhan dan penyakit. Saat ini diabetes melitus tidak hanya sebagai penyebab kematian prematur di seluruh dunia. Penyakit ini juga menjadi penyebab utama kebutaan, penyakit jantung dan gagal ginjal.

Menurut *World Health Organization* (WHO), diabetes melitus didefinisikan sebagai suatu penyakit kronis dengan multi etiologi yang ditandai oleh tingginya kadar gula darah disertai gangguan metabolisme karbohidrat, lipid dan protein yang diakibatkan karena insufisiensi fungsi insulin. Berdasarkan data WHO tahun 2016, 70% dari total kematian di dunia dan 90-95% merupakan kasus diabetes melitus (WHO, 2016). Banyak faktor yang memicu terjadinya diabetes melitus diantaranya tekanan darah sistolik, tekanan darah diastolik, lama waktu tidur, gaya bekerja, tingkat pengetahuan tentang diabetes, ukuran lingkar perut, riwayat keturunan diabetes, usia, kebiasaan berolahraga dan indeks massa tubuh. Faktor-faktor risiko ini seharusnya diperhatikan secara serius agar dapat mengurangi angka diabetes melitus akibat kadar gula darah yang meningkat. Oleh karena itu, pengklasifikasian perlu dilakukan terhadap masyarakat penderita diabetes dan non diabetes berdasarkan faktor-faktor yang mempengaruhinya.

Metode klasifikasi telah banyak digunakan dalam berbagai bidang, seperti bidang pendidikan, pemerintahan, kesehatan, teknologi, maupun sosial. Salah satu metode klasifikasi yang dapat digunakan adalah *Chi-Square Automatic Interaction*

Detection (CHAID). Metode CHAID merupakan suatu teknik iteratif yang menguji satu persatu variabel prediktor yang digunakan dalam klasifikasi dan menyusunnya berdasarkan pada tingkat signifikansi statistik *chi-square* terhadap variabel responnya (Gallagher dkk., 2000). Hasil dari pengklasifikasian dalam CHAID ditampilkan dalam sebuah diagram pohon.

Metode *Exhaustive Chi-Square Automatic Interaction Detection* merupakan pengembangan dari metode CHAID yang menjelaskan mengenai prosedur penyekatan dengan cara melihat seluruh kemungkinan penggabungan dari pasangan kategori secara bertahap (Soemartojo, 2000). Metode *Exhaustive* CHAID juga mengalami penyesuaian untuk koreksi *Bonferroni* pada jumlah kategori (Zhang dkk., 2014). Tahapan metode *Exhaustive* CHAID terdiri dari tiga tahap yaitu, tahap penggabungan, tahap pemisahan dan tahap penghentian. Hasil akhir dari tahapan tersebut adalah pohon klasifikasi yang menginterpretasikan variabel-variabel prediktor yang berpengaruh signifikan terhadap variabel responnya.

Penelitian sebelumnya yang telah menggunakan metode CHAID diantaranya yaitu Nada (2020) melakukan penelitian mengenai klasifikasi pengetahuan masyarakat terhadap kesiapsiagaan terjadinya gempa bumi. Tohari dkk. (2017) menggunakan metode CHAID dan *Exhaustive* CHAID pada klasifikasi produksi jagung di Pulau Jawa. Selain itu, Nurhasanah dkk. (2022) menerapkan metode CHAID dan *Exhaustive* CHAID pada studi kasus data masa studi mahasiswa Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Mulawarman.

Menelaah dari penelitian sebelumnya yang menggunakan metode CHAID dan *Exhaustive* CHAID untuk melakukan pengklasifikasian, maka peneliti tertarik untuk menerapkan metode *Exhaustive* CHAID sebagai pengembangan metode CHAID. Oleh karena itu, penelitian ini mengkaji tentang klasifikasi penderita diabetes dan non diabetes berdasarkan faktor yang mempengaruhinya menggunakan metode *Exhaustive* CHAID. Berdasarkan uraian tersebut, penulis mengajukan bahan skripsi dengan judul **“Penerapan Metode *Exhaustive Chi-Square Automatic Interaction Detection* pada Klasifikasi Penderita Diabetes dan Non Diabetes”**.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan untuk mendapatkan faktor-faktor yang berpengaruh signifikan terhadap penderita diabetes dan non diabetes menggunakan metode *Exhaustive Chi-Square Automatic Interaction Detection*, maka rumusan masalah yang mesti diselesaikan yaitu:

1. Bagaimana faktor-faktor yang berpengaruh signifikan terhadap penderita diabetes dan non diabetes berdasarkan hasil analisis metode *Exhaustive Chi-Square Automatic Interaction Detection*?
2. Bagaimana hasil klasifikasi penderita diabetes dan non diabetes dengan menggunakan metode *Exhaustive Chi-Square Automatic Interaction Detection*?

1.3 Batasan Masalah

Pembatasan masalah sangat diperlukan agar tidak terjadi penyimpangan dari tujuan penelitian sehingga pemecahan masalah akan lebih terfokus. Batasan masalah pada penelitian ini adalah:

1. Penelitian ini menggunakan data kadar gula darah, tekanan darah sistolik, tekanan darah diastolik, lama waktu tidur, gaya bekerja, tingkat pengetahuan tentang diabetes, ukuran lingkar perut, riwayat keturunan diabetes, usia, kebiasaan berolahraga dan indeks massa tubuh penderita diabetes dan non diabetes di Kabupaten Muna Barat.
2. Penelitian ini menggunakan metode *Exhaustive Chi-Square Automatic Interaction Detection* dalam pembentukan pohon klasifikasi yang dibatasi dengan penggunaan statistik uji *chi-square* dengan taraf signifikansi $\alpha = 5\%$.

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah diuraikan, adapun tujuan yang ingin dicapai pada penelitian ini adalah:

1. Mendapatkan faktor-faktor yang berpengaruh signifikan terhadap penderita diabetes dan non diabetes berdasarkan hasil analisis metode *Exhaustive Chi-Square Automatic Interaction Detection*.
2. Mendapatkan hasil klasifikasi penderita diabetes dan non diabetes dengan menggunakan metode *Exhaustive Chi-Square Automatic Interaction Detection*.

1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

1. Memberikan pengetahuan mengenai metode statistik yang dapat digunakan untuk suatu pengklasifikasian dengan menggunakan metode *Exhaustive Chi-Square Automatic Interaction Detection*.
2. Hasil dari penelitian ini diharapkan dapat menjadi referensi dan sumber informasi bagi masyarakat untuk mengetahui faktor-faktor yang memengaruhi penderita diabetes dan non diabetes.

BAB II

TINJAUAN PUSTAKA

2.1 Metode *Exhaustive* CHAID

Metode *Exhaustive Chi-Square Automatic Interaction Detection* adalah metode eksplorasi untuk mengklasifikasikan data dengan cara membangun pohon klasifikasi. Pohon klasifikasi pada metode *Exhaustive* CHAID dapat memberikan informasi berupa variabel prediktor yang berpengaruh secara signifikan terhadap variabel respon (Bagozzi, 1994). Metode *Exhaustive* CHAID merupakan pengembangan dari metode CHAID yang menjelaskan mengenai prosedur penyekatan dengan cara melihat seluruh kemungkinan penggabungan dari pasangan kategori secara bertahap (Soemartojo, 2000). Biggs dkk. (2006) mengemukakan bahwa metode *Exhaustive* CHAID merupakan evaluasi dari metode sebelumnya yaitu CHAID.

Metode *Exhaustive* CHAID menggunakan uji *chi-square* yang menguji satu persatu variabel prediktor kategori terhadap variabel respon kategori dalam klasifikasi (Gallagher dkk., 2000). Uji *chi-square* merupakan uji non parametrik yang sesuai untuk menguji hubungan antara variabel yang berbentuk kategori. Metode CHAID juga terdapat koreksi *Bonferroni* yang dilakukan ketika beberapa uji statistik untuk independensi dilakukan secara bersamaan. Namun untuk metode *Exhaustive* CHAID mengalami penyesuaian untuk koreksi *Bonferroni* pada jumlah kategori (Zhang dkk., 2014). Metode *Exhaustive* CHAID mengemukakan prosedur penyekatan dengan cara melihat seluruh kemungkinan penggabungan dari pasangan kategori, secara bertahap hingga tersisa dua kategori.

2.2 Variabel pada Metode *Exhaustive* CHAID

Variabel yang digunakan dalam metode *Exhaustive* CHAID terdiri atas variabel respon dan variabel prediktor. Metode klasifikasi dalam *Exhaustive* CHAID dilakukan berdasarkan pada hubungan yang ada antara kedua variabel tersebut. Menurut Gallagher dkk. (2000), variabel prediktor dalam metode *Exhaustive* CHAID dapat dibedakan menjadi tiga jenis. Variabel-variabel tersebut adalah sebagai berikut.

- Variabel monotonik yaitu variabel prediktor yang setiap kategori-kategori di dalamnya dapat digabungkan jika berurutan (data ordinal). Contohnya usia, tekanan darah dan pendapatan.
- Variabel bebas yaitu variabel prediktor yang setiap kategori-kategori di dalamnya dapat digabungkan meskipun tidak berurutan (data nominal). Contohnya pekerjaan dan jenis kelamin.
- Variabel mengambang yaitu variabel prediktor yang dapat diperlakukan sebagai variabel monotonik, kecuali untuk kategori yang *missing value*, yang dapat dikombinasikan dengan kategori manapun.

2.3 Uji Chi-Square

Statistik uji yang digunakan dalam metode *Exhaustive CHAID* adalah statistik uji *chi-square*. Statistik uji *chi-square* dapat digunakan untuk mengetahui independensi antara dua variabel. Misal suatu variabel pertama memiliki p kategori, yaitu $A_1, A_2, \dots, A_p$ dan variabel kedua memiliki q kategori, yaitu $B_1, B_2, \dots, B_q$. Sehingga struktur data uji *chi-square* dapat dilihat pada Tabel 2.1.

Tabel 2.1 Struktur Data Uji Chi-Square

Variabel Respon	Variabel Prediktor					Jumlah
	B_1	...	B_j	...	B_q	
A_1	n_{11}	...	n_{1j}	...	n_{1q}	$n_{1\cdot}$
...	...	...	...	...	...	...
A_i	n_{i1}	...	n_{ij}	...	n_{iq}	$n_{i\cdot}$
...	...	...	...	...	...	...
A_p	n_{p1}	...	n_{pj}	...	n_{pq}	$n_{p\cdot}$
Jumlah	$n_{\cdot 1}$	...	$n_{\cdot j}$	...	$n_{\cdot q}$	N

dengan

n_{ij} : Jumlah pengamatan A_i dan B_j , pada variabel respon ke $i = 1, \dots, p$, dan variabel prediktor ke $j = 1, \dots, q$

$n_{i\cdot}$: Jumlah pengamatan A_i , pada variabel respon ke $i = 1, \dots, p$

$n_{\cdot j}$: Jumlah pengamatan B_j , pada variabel prediktor ke $j = 1, \dots, q$

Hipotesis yang digunakan pada pengujian *chi-square* yaitu:

H_0 : Tidak terdapat hubungan antara variabel prediktor dan variabel respon

H_1 : Terdapat hubungan antara variabel prediktor dan variabel respon

Taraf signifikansi : α

Statistik Uji

$$\chi^2 = \sum_{i=1}^p \sum_{j=1}^q \frac{(n_{ij} - E_{ij})^2}{E_{ij}} \quad (2.1)$$

dengan

n_{ij} : Jumlah pengamatan A_i dan B_j , pada variabel respon ke $i = 1, \dots, p$, dan variabel prediktor ke $j = 1, \dots, q$

E_{ij} : Frekuensi harapan pengamatan baris ke- i dan kolom ke- j

Untuk menghitung frekuensi harapan masing-masing sel digunakan rumus sebagai berikut:

$$E_{ij} = \frac{n_{i.} \cdot n_{.j}}{n} \quad (2.2)$$

dengan

$n_{i.}$: Jumlah pengamatan A_i , variabel respon ke $i = 1, \dots, p$

$n_{.j}$: Jumlah pengamatan B_j , variabel prediktor ke $j = 1, \dots, q$

E_{ij} : Frekuensi harapan pengamatan baris ke- i dan kolom ke- j

n : Jumlah seluruh pengamatan

Kriteria pengambilan keputusan dalam uji *chi-square* yaitu H_0 ditolak jika $\chi^2_{hitung} > \chi^2_{\alpha; (p-1)(q-1)}$. Selain itu, pengambilan keputusan dapat juga dengan membandingkan *p-value* dengan taraf signifikansi (α). Jika *p-value* < taraf signifikansi (α) maka H_0 ditolak.

Statistik uji *chi-square* digunakan dalam dua cara pada analisis metode *Exhaustive CHAID*. Pertama, untuk menentukan apabila kategori-kategori dalam sebuah variabel prediktor bersifat seragam dan bisa digabungkan menjadi satu kategori. Kedua, ketika semua variabel prediktor sudah diringkas menjadi bentuk yang signifikan dan tidak mungkin digabung lagi, maka statistik uji *chi-square* digunakan untuk menentukan variabel prediktor mana yang paling signifikan untuk membagi kategori-kategori dalam variabel respon.

2.4 Koreksi Bonferroni

Koreksi *Bonferroni* adalah proses koreksi yang digunakan ketika beberapa uji statistik untuk kebebasan atau ketidakbebasan dilakukan secara bersamaan (Hasana & Kunto, 2006). Koreksi *Bonferroni* pada algoritma *Exhaustive CHAID* digunakan untuk penyesuaian *p-value* dari variabel prediktor yang telah terjadi pengurangan kategori. Penyesuaian ini dilakukan dengan melakukan pengalihan antara *p-value* sebelum terjadi pengurangan kategori dengan pengali *Bonferroni*. Misalkan jumlah kategori semula adalah c , dan jumlah kategori setelah terjadi penggabungan adalah r . Sehingga dalam metode *Exhaustive CHAID* penggabungan dilakukan bertahap sejak langkah ke $k = 1$ hingga tersisa dua kategori pada $k = c - 2$, maka:

$$N(c) = 1 + \sum_{k=1}^{c-2} N(c - k + 1, c - k) \quad (2.3)$$

untuk:

1) Variabel Monotonik

$$N(c - k + 1, c - k) = c - k \quad (2.4)$$

2) Variabel Bebas

$$N(c - k + 1, c - k) = \sum_{i=0}^{(c-k)-1} (-1)^i \frac{((c-k)-i)^{c-k+1}}{i!((c-k)-i)!} \quad (2.5)$$

3) Variabel Mengambang

Jika kategori mengambang tidak dapat digabungkan maka koefisien *Bonferroni* sama dengan penyesuaian *Bonferroni* pada skala ordinal.

$$N(c - 1) = 1 + \sum_{k=1}^{c-3} N(c - k + 1, c - k) \quad (2.6)$$

Jika kategori mengambang dapat digabungkan maka koefisien *Bonferroni* untuk tipe ini sama dengan koefisien *Bonferroni* tanpa kategori mengambang ($c - 1$ kategori), untuk skala ordinal ditambah jumlah gabungan bila kategori mengambang disertakan.

$$N = N(c - 1) + (c - 1) + 1 \quad (2.7)$$

dengan

N : Pengali *Bonferroni*

c : Jumlah kategori variabel prediktor awal

r : Jumlah kategori variabel prediktor setelah penggabungan

2.5 Algoritma Metode *Exhaustive* CHAID

Algoritma *Exhaustive* CHAID digunakan untuk menghasilkan pohon klasifikasi yang dapat digunakan untuk memprediksi nilai variabel respon. Secara garis besar algoritma ini dapat dibagi menjadi tiga tahapan yaitu tahap penggabungan, tahap pemisahan, dan tahap penghentian. Pohon klasifikasi *Exhaustive* CHAID diperoleh melalui tiga tahap tersebut, dimulai dari simpul akar dan dilakukan secara berulang pada setiap simpul yang terbentuk.

2.5.1 Tahap Penggabungan

Tahap pertama dalam algoritma *Exhaustive* CHAID adalah tahap penggabungan. Pada tahap ini diperiksa signifikansi dari masing-masing kategori variabel prediktor terhadap variabel respon. Kemudian kategori-kategori dengan nilai *p-value* yang mirip akan digabungkan. Tahap penggabungan untuk setiap kategori-kategori adalah sebagai berikut.

1. Membentuk tabel kontingensi dua arah untuk masing-masing kategori.
2. Menghitung statistik uji *chi-square* untuk setiap pasang kategori yang dapat dipilih untuk digabung menjadi satu.
3. Selanjutnya adalah menentukan nilai *p-value* berpasangan secara bersamaan, untuk masing-masing nilai *chi-square* yang berpasangan. Untuk pasangan kategori dengan *p-value* yang mirip, dilakukan penggabungan sehingga menjadi sebuah kategori tunggal.
4. Periksa kembali kesignifikanan kategori baru setelah digabung dengan kategori lainnya. Jika masih ada pasangan yang belum signifikan, maka ulangi langkah 3. Namun, jika semua signifikan maka lanjutkan langkah berikutnya.

2.5.2 Tahap Pemisahan

Tahap pemisahan dipergunakan untuk memilih kategori mana yang akan digunakan sebagai pemisah simpul terbaik. Pemisahan dilakukan dengan cara membandingkan nilai *p-value* yang diperoleh dari tahap penggabungan pada setiap kategori. Langkah tahap pemisahan adalah sebagai berikut.

1. Memilih variabel prediktor yang memiliki nilai *p-value* terkecil (paling signifikan) yang akan digunakan sebagai pemisah simpul.
2. Jika nilai *p-value* kurang dari atau sama dengan taraf signifikansi (α), maka variabel prediktor dilakukan pemisahan. Jika tidak ada variabel prediktor yang mempunyai nilai *p-value* signifikan maka lanjutkan ke tahap selanjutnya dan variabel prediktor tersebut menjadi simpul akhir.

2.5.3 Tahap Penghentian

Tahap penghentian dilakukan jika suatu proses pertumbuhan pohon harus dihentikan sesuai dengan peraturan penghentian di bawah ini.

1. Tidak ada variabel prediktor yang signifikan menunjukkan perbedaan terhadap variabel respon.
2. Jika pohon telah mencapai batas nilai maksimum, maka proses pertumbuhan pohon akan dihentikan. Misalkan telah ditentukan bahwa batas kedalaman pertumbuhan pohon klasifikasi adalah 3, maka ketika pertumbuhan pohon sudah mencapai kedalaman 3, pertumbuhan pohon klasifikasi dihentikan.

2.6 Matriks Konfusi

Sebuah metode klasifikasi diharapkan dapat melakukan klasifikasi semua data dengan benar, tetapi tidak dapat dipungkiri bahwa kinerja suatu metode tidak dapat 100% benar. Oleh karena itu, sebuah metode klasifikasi juga harus diukur kinerjanya untuk mengetahui bahwa hasil klasifikasi memberikan informasi yang akurat. Kinerja dari setiap metode klasifikasi dapat dievaluasi dengan menggunakan perhitungan statistik, yaitu keakuratan klasifikasi. Akurasi klasifikasi sebagai ukuran kualitas pengelompokan yang paling sederhana untuk mengevaluasi pengelompokan yang terkait dengan kebenaran dasar. Umumnya, pengukuran kinerja klasifikasi dapat dihitung dengan menggunakan matriks konfusi.

Matriks konfusi adalah tabel dengan empat kombinasi berbeda dari nilai prediksi dan nilai aktual. Ada empat istilah yang merupakan representasi hasil proses klasifikasi pada matriks konfusi yaitu *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN). Ketika kita melakukan uji, hasil prediksinya adalah *true* dan ternyata nilai aktualnya adalah *true*, maka disebut *true positive*.

Sedangkan ketika kita melakukan uji, hasil prediksinya adalah *false* dan ternyata nilai aktualnya adalah *false*, maka disebut *true negative*. Berikut ini merupakan sebuah matriks yang menunjukkan TP, TN, FP, dan FN, yang biasa dikenal dengan sebutan matriks konfusi.

Tabel 2.2 Matriks Konfusi

	P ^c (Prediksi)	N ^c (Prediksi)
P (Aktual)	<i>True Positive</i>	<i>False Negative</i>
N (Aktual)	<i>False Positive</i>	<i>True Negative</i>

Akurasi hasil prediksi dan *error rate* diperoleh dengan mengetahui jumlah data yang diklasifikasikan secara benar. Keakuratan klasifikasi dihitung dengan proporsi dari prediksi yang bernilai positif dan negatif. Cara menghitung akurasi hasil prediksi adalah sebagai berikut:

$$Akurasi = \frac{\text{banyaknya prediksi yang benar}}{\text{total banyaknya prediksi}} = \frac{TP+TN}{TP+FP+TN+FN} \quad (2.8)$$

Keterangan:

TP = *True Positive*

TN = *True Negative*

FP = *False Positive*

FN = *False Negative*

Selain itu, kinerja sebuah model dapat dinyatakan dalam bentuk *error rate* (kesalahan prediksi), yang diberikan oleh persamaan berikut:

$$Error\ rate = \frac{\text{banyaknya prediksi yang salah}}{\text{total banyaknya prediksi}} = \frac{FP+FN}{TP+FP+TN+FN} \quad (2.9)$$

Semua algoritma klasifikasi berusaha membentuk model yang mempunyai akurasi paling tinggi atau *error rate* yang rendah.

2.7 Diabetes Melitus

Diabetes melitus adalah penyakit kronis berupa gangguan metabolik yang ditandai dengan kadar gula darah yang melebihi batas normal. Diabetes melitus disebabkan oleh kenaikan kadar gula darah karena kerusakan sel beta pankreas sehingga produksi insulin tidak ada sama sekali. Insulin adalah hormon yang dihasilkan oleh pankreas untuk mencerna gula dalam darah. Diabetes adalah masalah

kesehatan masyarakat yang penting, menjadi salah satu dari empat penyakit tidak menular prioritas target tindak lanjut oleh para pemimpin dunia. Jumlah kasus dan prevalensi diabetes terus meningkat selama beberapa dekade terakhir (WHO, 2016).

Berdasarkan data WHO tahun 2016, 70% dari total kematian di dunia dan lebih dari setengah beban penyakit 90-95% merupakan kasus diabetes melitus. Secara global WHO memperkirakan 422 juta orang dewasa hidup dengan diabetes pada tahun 2014, dibandingkan dengan 108 juta pada tahun 1980. Organisasi *International Diabetes Federation* (IDF) memperkirakan sedikitnya terdapat 463 juta orang pada usia 20-79 tahun di dunia menderita diabetes pada tahun 2019 atau setara dengan angka prevalensi sebesar 9,3% dari total penduduk pada usia yang sama. Prevalensi diabetes diperkirakan meningkat seiring penambahan umur penduduk menjadi 19,9% atau 111,2 juta orang pada umur 65-79 tahun (Kemenkes, 2020).

Riset Kesehatan Dasar (Riskesdas) yang dilaksanakan pada tahun 2018 melakukan pengumpulan data penderita diabetes melitus pada penduduk yang berumur ≥ 15 tahun. Hasil Riskesdas 2018 menunjukkan bahwa prevalensi diabetes melitus di Indonesia berdasarkan diagnosis dokter pada umur ≥ 15 tahun sebesar 2%. Angka ini menunjukkan peningkatan dibandingkan prevalensi diabetes melitus pada penduduk ≥ 15 tahun pada hasil Riskesdas 2013 sebesar 1,5%. Namun prevalensi diabetes melitus menurut hasil pemeriksaan gula darah meningkat dari 6,9% pada 2013 menjadi 8,5% pada tahun 2018. Angka ini menunjukkan bahwa baru sekitar 25% penderita diabetes yang mengetahui bahwa dirinya menderita diabetes.