

DISERTASI

**MODEL DIAGNOSIS PENYAKIT DEGENERATIF
MENGUNAKAN
DECISION TREE ITERATIVE DICHOTOMISER 3
BERDASARKAN KOMBINASI KEANGGOTAAN FUZZY**



OLEH:


**Endang Sri Kresnawati
NIM 08013621001**

**PROGRAM STUDI S3 ILMU MIPA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN
ALAM UNIVERSITAS SRIWIJAYA
2024**

HALAMAN PENGESAHAN HASIL PENELITIAN DISERTASI
MODEL DIAGNOSIS PENYAKIT DEGENERATIF
MENGGUNAKAN
DECISION TREE ITERATIVE DICHOTOMISER 3
BERDASARKAN KOMBINASI KEANGGOTAAN FUZZY

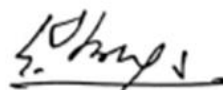
Oleh:
ENDANG SRI KRESNAWATI
08013681823005

Promotor



Prof. Yulia Resti, M.Si., Ph.D.
NIP. 197307191997022001

Inderalaya, Desember 2024
Co-Promotor



Dr. Bambang Supriatin, M.Si.
NIP. 197101261994121001

Mengetahui,
Koordinator Program Studi S3 Ilmu MIPA
Universitas Sriwijaya,



Prof. Drs. Dedi Rohendi, M.T., Ph.D.
NIP 196704191993031001

KATA PENGANTAR

Puji syukur ke hadirat Allah SWT, yang telah memberikan kemampuan yang tak terbatas, kesehatan yang tak terkira, dan keyakinan yang teguh hingga akhirnya disertasi ini dapat diselesaikan. Shalawat beserta salam senantiasa tercurahkan kepada baginda Nabi Muhammad SAW beserta keluarga, sahabat dan para pengikutnya hingga akhir zaman. Disertasi ini merupakan salah satu syarat untuk memperoleh gelar doktor Program Studi S3 Ilmu MIPA Universitas Sriwijaya untuk bidang mayor Matematika di Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Sriwijaya.

Disertasi ini mendeskripsikan tentang model *Decision Tree Iterative Dichotomiser 3 algorithm* berdasarkan diskritisasi menggunakan pendekatan fuzzy. Model yang dibangun berguna untuk memprediksi status Penyakit Jantung Koroner (PJK) dan Penyakit Diabetes Mellitus (PDM). Selesainya disertasi ini tentunya tidak lepas dari bantuan dan keterlibatan berbagai pihak, untuk itu pada kesempatan ini, penulis menyampaikan terimakasih kepada:

1. Prof. Hermasyah, Ph.D sebagai Dekan Fakultas MIPA Universitas Sriwijaya sekaligus sebagai Ketua Sidang Disertasi Penulis.
2. Prof. Dr. Dedy Rohendi, M.Si sebagai Koordinator Program Studi S3 Ilmu MIPA Universitas.
3. Prof. Yulia Resti, M.Si., Ph.D. sebagai Promotor Disertasi yang telah memotivasi akademik dan non akademik.
4. Dr. Bambang Suprihatin, M.Si sebagai Co-promotor, atas suportnya agar cepat menyelesaikan studi ini.

5. Kedua Penguji Sidang Disertasi, Ibu Dr. Yuli Andriani, M.Si. dan Ibu Dr. Novrikasari, M.Si yang telah memberikan perbaikan dalam penyelesaian disertasi ini.
6. Teman teman Mahasiswa Program Studi S3 MIPA yang selalu mendukung dalam penulisan penelitian disertasi ini. Serta teman teman seperjuangan dalam penyelesaian S3 nya masing masing.
7. Staf Fakultas MIPA dan Pascasarjana serta staf Prodi S3 Ilmu MIPA yang telah memberikan bantuan selama proses studi.
8. Terkhusus yang teramat berharga dalam hidup saya: Suami “Bapak Faizal, S.T.” dan anak-anak “Shalika Khairunisa-Naufal Athallah-Adifa Aderald Dary”.
Thank you so much for everything.
9. Kedua orang tuaku Bapak H. Tjik Deri dan Ibu Hj. Sri Banun Alamsyah, yang paling menginginkan penulis melanjutkan studi S3. Terima kasih untuk doa tulusnya.

Disertasi ini hanyalah setitik dari lautan ilmu yang luas. Masih banyak hal yang perlu dieksplorasi, dipelajari, dan dikembangkan. Semoga disertasi ini menjadi sumbangsih bagi perkembangan ilmu pengetahuan dan bermanfaat bagi kemaslahatan umat manusia.

Palembang, D e s e m b e r 2 0 2 4

Endang Sri Kresnawati
NIM. 08013681823005

SUMMARY

Degenerative diseases are one of the leading causes of chronic disability on a global scale, significantly affecting the quality of life of sufferers. These diseases also burden the health care system and individuals financially. The implementation of preventive strategies can be postponed until accurate prediction of disease status can be achieved. Degenerative diseases that are the main cause of death in many countries are coronary heart disease (CHD), while diabetes mellitus disease (DMD) increases the risk of CHD. Most of the predictor variables from the data set to predict the status of both diseases are continuous, but not all prediction methods can process continuous data, one of which is Decision Tree Iterative Dichotomiser3 (DTID3) method. This work aims to predict the status of both degenerative diseases, CHD and DMD using the DTID3 method with continuous type predictor variables transformed using discretization with the concept of set membership. Seven prediction models using the DTID3 method are proposed to predict the status of each degenerative disease. One DTID3 model uses the concept of crisp set membership, and six DTID3 models use the concept of fuzzy set membership (FDTID3). Each prediction model of FDTID3 represents one combination of fuzzy membership functions in discretizing continuous predictor variables, and one combination consists of three membership functions. The performance of the proposed FDTID3 model depends on the fuzzy membership functions used. The hypothesis that the performance of the seven proposed models differs at least in one metric and that the performance of the FDTID3 models is higher than the DTID3 model discretized using the concept of crisp sets has been proven.

Keywords: DTID3; Degenerative Disease Status; Discretization; Fuzzy Membership Function; Model Performance

RINGKASAN

Penyakit degeneratif merupakan salah satu penyebab utama kecacatan kronis dalam skala global dan sangat mempengaruhi kualitas hidup penderitanya. Penyakit-penyakit ini juga membebani sistem layanan kesehatan dan individu secara finansial. Penerapan strategi pencegahan dapat ditunda hingga prediksi status penyakit yang akurat dapat dicapai. Penyakit degeneratif yang menjadi penyebab utama kematian di banyak negara adalah Penyakit Jantung Koroner (PJK), sedangkan Penyakit Diabetes Melitus (PDM) meningkatkan risiko PJK. Sebagian besar variabel prediktor dari kumpulan data untuk memprediksi status kedua penyakit tersebut bersifat kontinu, namun tidak semua metode prediksi dapat mengolah data kontinu, salah satunya adalah metode *Decision Tree Iterative Dichotomiser3* (DTID3). Karya ini bertujuan untuk memprediksi status penyakit baik degeneratif, PJK maupun PDM menggunakan metode DTID3 dengan variabel prediktor tipe kontinu ditransformasikan menggunakan diskritisasi dengan konsep keanggotaan himpunan. Tujuh model prediksi dengan metode DTID3 diusulkan untuk memprediksi status masing-masing penyakit degeneratif. Satu model DTID3 menggunakan konsep keanggotaan himpunan tajam, dan enam model DTID3 menggunakan konsep keanggotaan himpunan fuzzy (FDTID3). Setiap model prediksi FDTID3 mewakili satu kombinasi fungsi keanggotaan fuzzy dalam mendiskritisasi variabel prediktor kontinu, dan satu kombinasi terdiri dari tiga fungsi keanggotaan. Kinerja model FDTID3 yang diusulkan bergantung pada fungsi keanggotaan fuzzy yang digunakan. Hipotesis bahwa performa ketujuh model yang diusulkan berbeda setidaknya dalam satu metrik dan performa model FDTID3 lebih tinggi dibandingkan model DTID3 yang didiskritisasi menggunakan konsep himpunan tegas telah terbukti.

Kata Kunci: DTID3, Status Penyakit Degeneratif, Diskritisasi, Fungsi Keanggotaan Fuzzy, Performa Model

HALAMAN PERNYATAAN INTEGRITAS

Yang bertanda tangan di bawah ini:

Nama : Endang Sri Kresnawati

NIM : 08013681823005

Judul : Model Diagnosis Penyakit Degeneratif Menggunakan *Decision Tree Iterative Dichotomiser 3* Berdasarkan Kombinasi Keanggotaan Fuzzy

Menyatakan bahwa Disertasi saya merupakan hasil kerja saya sendiri didampingi oleh Tim Promotor dan Co-Promotor serta bukan hasil penjiplakan/plagiat. Apabila ditemukan unsur penjiplakan/plagiat dalam Disertasi ini, maka saya bersedia menerima sanksi akademik dari Universitas Sriwijaya sesuai dengan peraturan yang berlaku.

Demikian pernyataan ini saya buat dalam keadaan sadar tanpa ada paksaan dari siapapun.



g, Desember 2024

Endang Sri Kresnawati
NIM. 08013681823005

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PENGESAHAN	ii
SUMMARY	v
RINGKASAN	vi
HALAMAN PERNYATAAN INTEGRITAS	vii
DAFTAR ISI	viii
DAFTAR GAMBAR	x
DAFTAR TABEL	xii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	7
1.3 Batasan Masalah	8
1.4 Tujuan Penelitian	8
1.5 Manfaat Penelitian	8
1.6 Keterbaruan.....	8
1.7 Kerangka Berpikir Penelitian.....	11
BAB II TINJAUAN PUSTAKA	12
2.1 Diskriminan Analisis.....	12
2.2 Naive Bayes.....	15
2.3 Decision Tree.....	16
2.4 Fuzzy Decision Tree	17
2.5 Validasi Silang <i>5 Fold</i>	23

2.6	Metrik Evaluasi Prediksi.....	24
2.7	Penyakit Degeneratif.....	25
2.8	Unsur-unsur Diagnosis Penyakit.....	27
BAB III METODOLOGI PENELITIAN		
3.1.	Tempat Penelitian.....	34
3.2.	Waktu Penelitian.....	34
3.3.	Alat Penelitian.....	34
3.4.	Data Set Penelitian.....	34
3.5.	Metode Penelitian.....	37
BAB IV HASIL DAN PEMBAHASAN.....		44
4.1.	Analisis Diskriminan.....	44
4.2.	Multiple naive bayes	49
4.3.	Decision Tree.....	57
4.4.	Fuzzy Decision Tree.....	61
BAB V KESIMPULAN DAN SARAN.....		88
5.1	Kesimpulan.....	117
5.2	Saran.....	118
DAFTAR PUSTAKA.....		118

DAFTAR GAMBAR

Gambar 1.1 Kerangka Berfikir Penelitian	11
Gambar 2.1. Validasi Silang 5 Fold.....	23
Gambar 3.1 Bagan Alur Pemodelan Diskriminan Analisis.....	30
Gambar 3.2 Bagan Alur Pemodelan Naive Bayes dan Decision Tree.....	32
Gambar 3.2 Bagan Alur Pemodelan Fuzzy Decision Tree.....	35
Gambar 4.1 Mean and Standar Deviation of Explanatory Variable.....	36
Gambar 4.2 Sebaran Data Latih PJK.....	38
Gambar 4.3 Komposisi Data Uji dan Data Latih 10 Fold.....	44
Gambar 4.4 Struktur Pohon pada Node 1.....	56
Gambar 4.5 Struktur Lengkap Pohon Keputusan Dataset PJK.....	34
Gambar 4.6 Iterasi 1 Model FDTID3-4 dengan Validasi Lengkap untuk Status PJK..	44
Gambar 4.7 Struktur Lengkap Decision Tree.....	60
Gambar 4.8 Proses Perhitungan Akurasi.....	60
Gambar 4.9 Matriks Konfusi Iterasi Satu.....	61
Gambar 4.10 Histogram dan Diagram Batang Variabel Prediktor PJK.....	62
Gambar 4.11 Ringkasan Variabel Kategori Dataset PJK.....	64
Gambar 4.12 Iterasi Pertama FDTID3 PJK.....	76
Gambar 4.13 Kinerja FDTID3-1 PJK Berdasarkan Validasi Silang 5 Fold.....	76
Gambar 4.14 Kinerja FDTID3-2 PJK Berdasarkan Validasi Silang 5 Fold.....	78
Gambar 4.15 Kinerja FDTID3-3 PJK Berdasarkan Validasi Silang 5 Fold.....	78
Gambar 4.16 Kinerja FDTID3-4 PJK Berdasarkan Validasi Silang 5 old.....	79

Gambar 4.17	Kinerja FDTID3-5 PJK Berdasarkan Validasi Silang 5 Fold.....	79
Gambar 4.18	Kinerja FDTID3-6 PJK Berdasarkan Validasi Silang 5 Fold.....	80
Gambar 4.19	Histogram dan Bar Plot Variabel Prediktor dalam Dataset PDM.....	85
Gambar 4.20	Struktur Pohon Keputusan Akhir Iterasi Satu Dataset PDM.....	89
Gambar 4.21	Kinerja Enam Model Prediksi PDM FDTID3 Berdasarkan Validasi Silang 5 Fold.....	106

DAFTAR TAB EL

Tabel 3.1 Variabel Prediktor Dataset PJK.....	26
Tabel 3.2 Variabel Prediktor Dataset PDM.....	28
Tabel 4.1 Komposisi Data Uji dan Data Latih.....	36
Tabel 4.2 Uji Gaussian Multivariate Data latih.....	37
Tabel 4.3 Uji Homogenitas.....	39
Tabel 4.4 Uji Kesamaan Vektor Rata-rata Menggunakan Pillai Trace.....	39
Tabel 4.5 Parameter ADK.....	39
Tabel 4.6 Validasi Kinerja Model.....	40
Tabel 4.7 Diskritisasi Variabel Kontinu Dataset PJK.....	44
Tabel 4.8 Komposisi Data Uji dan Data Latih.....	44
Tabel 4.9 Diskritisasi Tegas Berdasarkan Informasi Sebelumnya tentang PJK.....	45
Tabel 4.10 Interval Diskritisasi Dataset PJK.....	45
Tabel 4.11 Hasil Diskritisasi Variabel prediktor PJK.....	46
Tabel 4.12 Variabel Prediktor dan Kategorinya.....	46
Tabel 4.13 Interval diskritisasi Dataset PDM.....	47
Tabel 4.14 Persentase Data Latih dan Data Uji.....	47
Tabel 4.15 Mean and Standard Deviation.....	48
Tabel 4.16 Hasil Diskritisasi Dataset PDM.....	48
Tabel 4.17 Model Pembelajaran Multinomial Naive Bayes.....	49
Tabel 4.18 Evaluasi Hasil Prediksi Dataset PJK.....	50

Tabel 4.19 Evaluasi Hasil Prediksi Dataset PDM.....	51
Tabel 4.20 Komposisi data Uji dan Data Latih.....	53
Tabel 4.21 Hasil Diskritisasi Variabel Prediktor PJK.....	54
Tabel 4.22 Rekap Perhitungan Entropi Gain Node 1 PJK.....	55
Tabel 4.23 Matriks Konfusi PJK Menggunakan Decision Tree.....	56
Tabel 4.24 Evaluasi Hasil Prediksi PJK Menggunakan Decision Tree.....	57
Tabel 4.25 Evaluasi Hasil Prediksi PJK	58
Tabel 4.26 Hasil Diskritisasi Dataset PDM.....	58
Tabel 4.27 Hasil Evaluasi Kinerja Model Decision tree PDM.....	61
Tabel 4.28 Ringkasan Variabel Kontinu dalam Dataset PJK.....	63
Tabel 4.29 Interval Diskritisasi Fuzzy Dataset PJK.....	65
Tabel 4.30 Komposisi Data pada setiap Fold PJK.....	66
Tabel 4.31 Matriks Konfusi Prediksi Status PJK Iterasi 1 FDTI3 dengan 13 Variabel.....	76
Tabel 4.32 Kinerja Prediksi Status PJK.....	81
Tabel 4.33 ANOVA Model yang Diusulkan PJK.....	82
Tabel 4.34 Uji Tukey-Cramer PJK.....	83
Tabel 4.35 Ringkasan Variabel Kontinu dalam Dataset PM	86
Tabel 4.36 Komposisi Data pada Setiap Fold PDM.....	88
Tabel 4.37 Matriks Konfusi Prediksi Status PJK Iterasi 1 FDTI3 dengan 13 Variabel.....	105
Tabel 4.39 Kinerja Prediksi Status PJK.....	107
Tabel 4.40 ANOVA Model yang Diusulkan PDM.....	108
Tabel 4.41 Uji Tukey-Cramer PDM.....	109

Tabel 4.41 Perbandingan Kinerja Model Status PJK.....	114
Tabel 4.42 Perbandingan Kinerja Model Status PDM.....	115

BAB I

PENDAHULUAN

1.1 Latar Belakang

Variabel tipe kontinu dapat ditemukan di banyak kasus kehidupan nyata, terutama kumpulan data di bidang medis (Kuo *et al.*, 2023), (Nezhad, Zahedi and Farahani, 2022). Dataset ini berguna untuk memprediksi status penyakit, termasuk penyakit degeneratif (Kee *et al.*, 2023), (Abdalrada *et al.*, 2022). Penyakit degeneratif terjadi akibat menurunnya fungsi organ dan jaringan tubuh secara perlahan dan dapat menyerang saraf, tulang belakang, persendian, dan otak (Eadie, 1974). Penyakit ini cenderung memburuk seiring berjalannya waktu dan berdampak pada kualitas hidup penderitanya (Batista and Pereira, 2016), (Harahap, Andayani and Utara, 2015), (Barendregt, 1998). Penyakit degeneratif bersifat kronis karena tidak menular, berkembang lambat, dan berlangsung lama (Renzo *et al.*, 2023). Selain itu, penyakit ini menduduki peringkat penyebab utama kecacatan kronis dalam skala global. Penyakit degeneratif berdampak pada lebih dari 30% populasi di seluruh dunia, dan 70% sumber daya kesehatan masyarakat dialokasikan untuk pengobatan penyakit tersebut. Selain itu, penyakit degeneratif memberikan tekanan finansial yang besar pada sistem layanan kesehatan dan individu (Livingston *et al.*, 2021).

Banyak kematian yang disebabkan oleh Penyakit Jantung Koroner (PJK), suatu kondisi degeneratif yang disebabkan oleh penyumbatan atau penyempitan arteri koroner akibat timbunan lemak. Selain itu, penyakit ini telah menjadi

penyebab utama kematian dalam skala global (Nelwan *et al.*, 2016). Peningkatan substansial PJK yang tercatat dari 12,1 juta kematian pada tahun 1990 menjadi 20,5 juta kematian pada tahun 2021. Jumlah ini mewakili sekitar 33,33 persen dari seluruh kematian global (Mariachiara Di Cesare *et al.*, 2023). Kematian dini sebagian besar disebabkan oleh penyakit kardiovaskular. Seratus empat puluh enam negara melaporkan kematian laki-laki dan 98 negara melaporkan kematian perempuan (Mariachiara Di Cesare *et al.*, 2023). Menurut Perkiraan Kesehatan Global yang diterbitkan oleh Organisasi Kesehatan Dunia (WHO) pada tahun 2019, diabetes bertanggung jawab atas sekitar 1,5 juta kematian (Antini *et al.*, 2024). Penyakit Diabetes Melitus (PDM) merupakan salah satu penyakit degeneratif yang berpotensi menimbulkan berbagai penyakit berat (WHO, 2016). Jumlah penderita Diabetes Mellitus (DM) diperkirakan mencapai 313 juta pada tahun 2040, menurut WHO (IDF, 2017). Tren ini diperkirakan akan terus memburuk seiring berjalannya waktu. Mengenai tekanan ekonomi, biaya pengobatan PDM akan berlipat ganda dari \$13.700 per tahun pada tahun 2030 (Abdollahi *et al.*, 2019). Strategi pencegahan dapat diterapkan pada tahap awal jika prediksi status diabetes melitus dan penyakit jantung koroner secara akurat dapat dilakukan dan pengurangan biaya yang besar dapat dicapai dalam pengobatan kedua penyakit tersebut pada (Cavan, Makaroff and Fernandes, 2016), (WHO, 2019), (Hossen, 2022).

Beberapa penelitian telah melakukan prediksi dengan menggunakan berbagai metode mengenai keberadaan kedua penyakit degeneratif tersebut dan mengusulkan beberapa pendekatan untuk meningkatkan kinerjanya, seperti teknik

ensemble (Chowdary, 2020), (Hassan *et al.*, 2022), penyeimbangan kelas pada dataset (Hossen, 2022), (Tasin *et al.*, 2023), fitur penskalaan (Patil and Bhosale, 2023), memilih variabel signifikan (Kanwal *et al.*, 2022), (Selvarathi and Varadhaganapathy, 2023), (Karthikeyini *et al.*, 2023), menangani nilai yang hilang (Palanivinayagam and Damaševičius, 2023), (Benarbia, 2022), atau melakukan prapemrosesan sebelum data diprediksi, seperti mengubah data menjadi tipe tertentu karena metode prediksi memerlukan variabel prediktor bertipe tertentu (Dougherty, Kohavi and Sahami, 1995).

Decision Tree Iterative Dichotomiser3 (DTID3) merupakan metode prediksi nonparametrik yang seringkali memberikan kinerja prediksi yang memuaskan dalam banyak kasus, namun variabel prediktor dalam metode ini harus bersifat kategorikal. Dalam hal variabel prediktor numerik perlu ditransformasikan terlebih dahulu ke dalam tipe kategorikal, dan salah satu teknik transformasinya adalah diskritisasi tegas yang dikenal dengan diskritisasi (saja) (García, Luengo and Herrera, 2015). Diskritisasi juga dapat memperluas pengetahuan mengenai tipe data kontinu (Roy and Pal, 2003) dan meningkatkan kinerja model (Resti, 2024), (Kresnawati *et al.*, 2021). Namun demikian, ambiguitas dapat terjadi akibat diskritisasi (Resti, Irsan, Amini, *et al.*, 2022). Konsep keanggotaan himpunan fuzzy dapat digunakan untuk memperbaiki ambiguitas dalam konteks diskritisasi dan dikenal sebagai diskritisasi fuzzy (Resti *et al.*, 2023), (Femina and Sudheep, 2020), (Shanmugapriya *et al.*, 2017).

Penggunaan diskritisasi fuzzy dalam beberapa metode prediksi telah terbukti meningkatkan kinerja prediksi, seperti metode Naïve Bayes (Resti *et al.*,

2023), (Femina and Sudheep, 2020), (Tutuncu and Kayaalp, 2019), fungsi basis radial jaringan saraf (NNRBF) (Algehyne *et al.*, 2022), *Multilayer Perceptron* (MLP) (Roy and Pal, 2003), dan DTID3 (Resti, Irsan, Amini, *et al.*, 2022), (Altay and Cinar, 2016). Beberapa kombinasi fungsi keanggotaan yang berbeda telah diterapkan dalam studi ini, serta aturan pemilihan keanggotaan akhir. Sayangnya hingga saat ini belum ada pedoman dalam pemilihan kombinasi fungsi keanggotaan fuzzy dalam mendiskritisasi suatu variabel prediktor serta aturan pemilihan keanggotaan akhir, sehingga *trial and error* (Altay and Cinar, 2016), (Araniba, 1994) masih menjadi solusi terbaik dalam menentukan kinerja model prediksi (Resti *et al.*, 2023). Mengenai besaran diskritisasi variabel prediktor mengacu pada justifikasi ahli (Tutuncu and Kayaalp, 2019), (Kresnawati *et al.*, 2021) atau pengetahuan sebelumnya (Resti *et al.*, 2023).

Perbedaan kombinasi fungsi keanggotaan fuzzy dapat mempengaruhi performa model prediksi. Penelitian (Resti *et al.*, 2023) mengklasifikasikan penyakit dan hama tanaman jagung ke dalam tujuh kelas dengan menggunakan metode Naive Bayes dan enam kombinasi fungsi keanggotaan fuzzy yang terdiri dari 4 kombinasi sesama fungsi linier dan dua kombinasi fungsi nonlinier dan fungsi linier. Penelitian tersebut memperoleh performa model terbaik dengan menggunakan fungsi linier sesama: linier menurun-meningkat dan segitiga. Penelitian (Resti, Irsan, Amini, *et al.*, 2022) yang mengklasifikasikan penyakit dan hama tanaman jagung ke dalam enam kelas dengan menggunakan metode DTID3 dan mendiskritisasi variabel prediktor dengan menggabungkan fungsi keanggotaan fuzzy nonlinier dan linier juga menunjukkan peningkatan kinerja model

menggunakan diskritisasi tegas. Demikian pula penelitian (Resti *et al.*, 2020) juga menunjukkan peningkatan kinerja metode asli (tanpa diskritisasi fuzzy). Penelitian mengklasifikasikan tipe kaleng menggunakan metode Naive Bayes dan mendiskritisasi variabel prediktor dengan kombinasi fungsi keanggotaan fuzzy sesama linier yaitu linier *decreasing-increasing* dan segitiga. Ketiga penelitian tersebut menggunakan aturan pemilihan keanggotaan akhir yang bernilai maksimal. Peningkatan kinerja setiap penelitian adalah akurasi sebesar 0,7 dan *recall* sebesar 3,95 (Resti *et al.*, 2023) akurasi sebesar 3,23 dan *recall* sebesar 11,8 (Resti, Irsan, Amini, *et al.*, 2022) akurasi sebesar 34,93 dan *recall* sebesar 35,08 (Resti *et al.*, 2020).

Aturan seleksi lainnya adalah *arithmetic mean* seperti pada penelitian (Femina and Sudheep, 2020) dan (Fernandez, Ito and Cruz-piris, 2022) yang memprediksi status PJK. Keduanya menggunakan metode prediksi Naive Bayes dan menggunakan fungsi keanggotaan fuzzy seragam untuk diskritisasi yaitu semua segitiga dan semua trapesium dengan peningkatan kinerja akurasi sebesar 4.03, *recall* sebesar 4.05 (Femina and Sudheep, 2020) dan akurasi sebesar 7.5 (Tutuncu and Kayaalp, 2019).

Selain dua aturan pemilihan keanggotaan akhir, beberapa penelitian menggunakan defuzzifikasi dengan metode tertentu seperti *centroid of area* (COA) (Algehyne *et al.*, 2022), *mean of maksimum* (MoM), (Shanmugapriya *et al.*, 2017), atau *centroid of area* (COA) (Fernandez, Ito and Cruz-piris, 2022). Defuzzikasi merupakan bagian dari aturan fuzzy dalam logika fuzzy. Dalam tugas prediksi, metode logika fuzzy dapat berdiri sendiri seperti pada penelitian (Fernandez, Ito

and Cruz-piris, 2022) atau diintegrasikan dengan metode prediksi lain seperti jaringan saraf (Shanmugapriya *et al.*, 2017), (Algehyne *et al.*, 2022). Penelitian (Algehyne *et al.*, 2022) yang menggunakan fungsi keanggotaan fuzzy seragam (semua trapesium) untuk memprediksi kanker payudara memperoleh peningkatan akurasi sebesar 3,72 dan *recall* sebesar 10,96. Namun pada penelitian (Shanmugapriya *et al.*, 2017) yang juga menggunakan fungsi keanggotaan fuzzy yang seragam (semua segitiga) memperoleh kinerja yang tidak terduga dalam memprediksi status PDM, di mana kinerja model awal tidak meningkat, bahkan menurun (akurasi sebesar 18,35 dan recall sebesar 23,97).

Hal ini diduga disebabkan banyaknya diskritisasi pada masing-masing variabel prediktor tidak mengacu pada justifikasi ahli padahal data yang digunakan adalah data medis yaitu dataset PDM. Selain itu, metode prediksi *Neural Network* tidak memerlukan transformasi variabel prediktor menjadi kategorikal karena dapat langsung mengolah data bertipe numerik diskritisasi fuzzy. Pada dataset PJK, penelitian (Femina and Sudheep, 2020) juga secara subyektif membedakan variabel prediktor menjadi dua kategori, bukan berdasarkan justifikasi ahli, yang dapat berbeda-beda untuk setiap variabel. Dari seluruh penelitian tersebut, belum ditemukan kombinasi fungsi keanggotaan fuzzy non-linier dalam mendiskritisasi variabel prediktor. Khusus penelitian yang membahas mengenai prediksi status penyakit PJK dan PDM, belum ditemukan penelitian yang menggunakan nilai maksimum sebagai aturan pemilihan keanggotaan akhir, serta kombinasi fungsi keanggotaan fuzzy yang berbeda.

Oleh karena itu, kontribusi utama penelitian ini adalah membangun model

prediksi status kedua penyakit degeneratif, PJK dan PDM, menggunakan metode metode DTID3 dengan mendiskritisasi variabel prediktor tipe kontinu menggunakan konsep keanggotaan himpunan fuzzy. Satu model DTID3 dibangun dengan menggunakan konsep keanggotaan himpunan tegas, dan enam model DTID3 yang variabel-variabelnya didiskritisasi menggunakan konsep keanggotaan himpunan fuzzy, disebut model *Fuzzy Decision Tree ID3* (FDTID3). Keenam kombinasi tersebut terdiri dari dua kombinasi fungsi linier sesama (linier menurun-naik dan segitiga, linier menurun-naik dan trapesium), dua kombinasi fungsi linier dan nonlinier (sigmoid dan segitiga menurun-naik, sigmoid menurun-naik dan trapesium), dan dua kombinasi fungsi nonlinier lainnya (sigmoid dan beta menurun-meningkat, sigmoid menurun-meningkat dan pi). Aturan pemilihan keanggotaan akhir yang digunakan adalah nilai maksimal. Penelitian ini juga memberikan hipotesis bahwa performa ketujuh model yang dibangun berbeda setidaknya pada satu metrik. Performa keenam model FDTID3 lebih tinggi dibandingkan model DTID3 yang dibangun menggunakan konsep keanggotaan himpunan tegas.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan di atas dapat dirumuskan masalah dalam penelitian ini, yaitu bagaimana memprediksi status penyakit degeneratif menggunakan model *Fuzzy Decision Tree ID3* (DTID3) dengan diskritisasi menggunakan fungsi kombinasi keanggotaan fuzzy. Selanjutnya bagaimana kinerja model yang dihasilkan, kombinasi mana yang menghasilkan kinerja terbaik.

1.3 Batasan Masalah

Batasan masalah dalam penelitian ini yaitu: Algoritma pohon keputusan menggunakan ID3, sehingga variabel non kategorik didiskritisasi menggunakan kombinasi fungsi keanggotaan fuzzy.

1.4 Tujuan Penelitian

Penelitian ini bertujuan membangun model diagnosis penyakit degeneratif, khususnya PJK dan PDM menggunakan Fuzzy *Decision Tree* dengan algoritma ID3. Variabel prediktor yang berupa data numerik didiskritisasi menggunakan kombinasi beberapa fungsi keanggotaan fuzzy. Model diagnosis ini digunakan untuk memprediksi status penyakit degeneratif tersebut.

1.5 Manfaat Penelitian

Manfaat penelitian ini adalah mendapatkan hasil prediksi status penyakit degeneratif, yaitu status PJK dan PDM berdasarkan kombinasi keanggotaan fuzzy. Selain itu penelitian ini mendapatkan kombinasi fungsi keanggotaan fuzzy yang menghasilkan model dengan kinerja terbaik.

1.6 Keterbaruan

Diskritisasi adalah teknik yang digunakan untuk merubah variabel prediktor bertipe numerik ke tipe kategorik dipelajari oleh model DTID3. Teknik ini dapat menimbulkan ambiguitas. Hal ini dapat diatasi dengan menerapkan fuzzy yang merupakan pendekatan matematis untuk menjelaskan variabel yang bersifat ambigu, kabur, atau tidak jelas. Diskritisasi yang sudah menerapkan konsep

keanggotaan fuzzy disebut diskritisasi fuzzy. Di sisi lain, konsep keanggotaan fuzzy dideskripsikan melalui fungsi keanggotaan fuzzy. Banyak jenis fungsi keanggotaan fuzzy, baik linier maupun non linier. Hasil pengkategorian variabel numerik sangat tergantung pada pemilihan fungsi keanggotaan yang dipakai. Sehingga berdampak pada performa model prediksi. Diskritisasi Fuzzy telah banyak digunakan peneliti sebelumnya.

Femina dan Sudheep (2020) menggunakan fungsi keanggotaan fuzzy seragam pada kategori diskritisasi ,yaitu semua **segitiga** dan semua **trapesium** dengan metode Naive Bayes. Kasusnya dataset penyakit jantung, di mana pengkategorian variabel prediktor bukan berdasarkan justifikasi ahli.

Penelitian mengklasifikasikan tipe kaleng menggunakan metode Naive Bayes yang dilakukan (Resti *et al.*, 2020) mendiskritisasi variabel prediktor dengan kombinasi fungsi keanggotaan fuzzy sesama linier yaitu **linier menurun**, **linier menaik**, dan **segitiga**. Pengkategorian variabel prediktor berdasarkan justifikasi ahli.

Fernandez, Ito, dan Cruz-piriz (2022) meneliti kasus perilaku pengendara mobil menggunakan sistem fuzzy *rule based*, dengan menerapkan fungsi keanggotaan fuzzy seragam yaitu semua **segitiga** dan semua **trapesium** dalam proses diskritisasi.

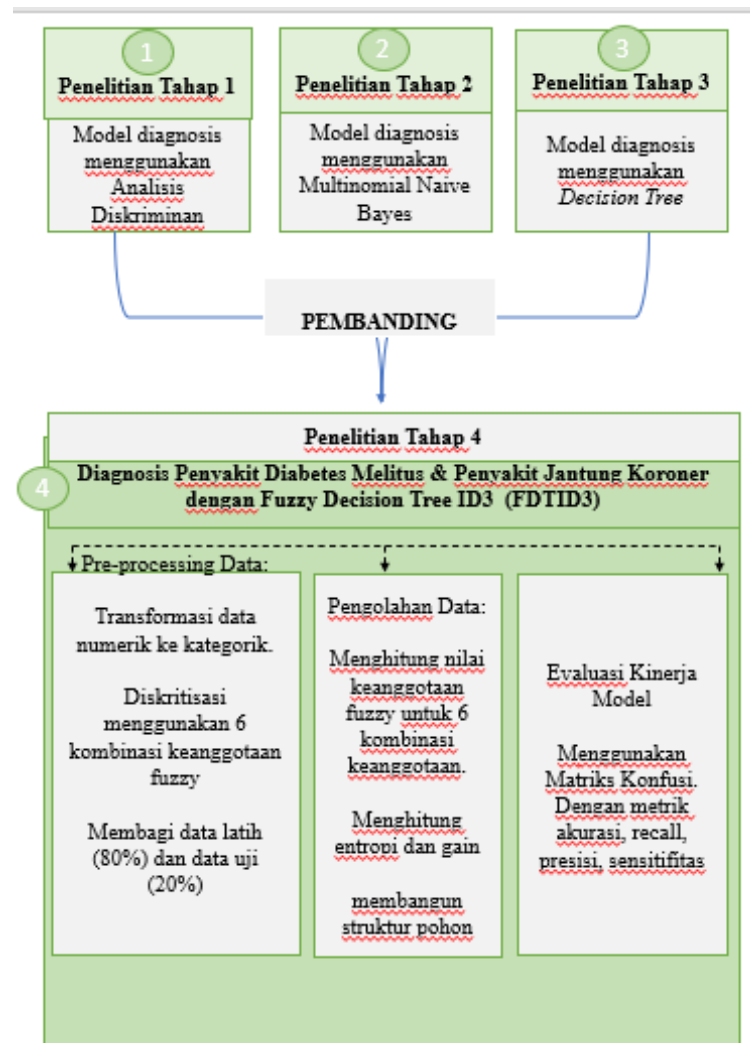
Penelitian Resti *et al* (2022) pada kasus kasus penyakit dan hama tanaman jagung, memilih fungsi keanggotaan fuzzy nonlinier dan linier (**sigmoid menurun-sigmoid meningkat** dan **segitiga**) pada metode *Decision Tree* ID3.

Selanjutnya pada penelitian (Resti *et al.*, 2023), menerapkan enam

kombinasi fungsi keanggotaan fuzzy yang terdiri dari 4 kombinasi sesama fungsi linier (semua **segitiga**, semua **trapesium**, **linier menurun-meningkat** dan **segitiga, menurun- meningkat linier** dan **trapesium**) dan dua kombinasi fungsi nonlinier dan fungsi linier (**sigmoid menurun-meningkat** dan **segitiga, menurun-meningkat sigmoid** dan **trapesium**) pada diskritisasi metode Naive Bayes untuk kasus penyakit dan hama tanaman jagung. Pengkategorian variabel prediktor berdasarkan justifikasi ahli.

Oleh karena itu, kebaruan penelitian ini adalah membangun model prediksi status kedua penyakit degeneratif, PJK dan PDM, menggunakan metode *Decision Tree ID3* (DTID3). Metode DTID3 mendiskritisasi variabel prediktor tipe kontinu menggunakan konsep keanggotaan fuzzy dengan fungsi dan kombinasi fungsi keanggotaan yang berbeda. Satu model DTID3 dibangun dengan menggunakan konsep keanggotaan himpunan tegas, dan enam model menggunakan enam kombinasi fungsi keanggotaan fuzzy. Dua kombinasi fungsi linier sesama (**linier menurun-naik** dan **segitiga, linier menurun-naik dan trapesium**), dua kombinasi fungsi linier dan nonlinier (**sigmoid** dan **segitiga menurun-naik, sigmoid menurun-naik dan trapesium**), dan dua kombinasi fungsi nonlinier lainnya (**sigmoid menurun-meningkat dan beta, sigmoid menurun-meningkat dan pi**). Aturan pemilihan keanggotaan akhir yang digunakan adalah nilai maksimal nilai keanggotaan. Selain itu penentuan jumlah kategori pada variabel prediktor merujuk pada referensi valid (ahli).

1.7. Kerangka Berfikir Penelitian



Gambar 1. Kerangka Berpikir Penelitian

DAFTAR PUSTAKA

- Abdalrada, A. S. *et al.* (2022). Machine learning models for prediction of co-occurrence of diabetes and cardiovascular diseases : a retrospective cohort study. *Journal of Diabetes & Metabolic Disorders*, 21, 251–261. <https://doi.org/10.1007/s40200-021-00968-z>
- Abdollahi, J. *et al.* (2019). Improving diabetes diagnosis in smart health using genetic-based ensemble learning algorithm approach to iot infrastructure. *Future Generation in Distributed Systems Journal*, 1(2), 26–33.
- Aggarwal, H., Arora, H. D. dan Kumar, V. (2020). Constructing a data mining model using fuzzy decision tree. *International Journal of Advanced Science and Technology*, 29(3), 3924–3934.
- Alcala-Fdez, J. *et al.* (2011). Keel data-mining software tool: data set repository, integration of algorithms and experimental analysis framework. *Journal of Multiple-Valued Logic and Soft Computing*, 17(2-3), 255–287.
- Alcalá-Fdez, J. *et al.* (2009). Keel: a software tool to assess evolutionary algorithms for data mining problems. *Soft Computing*, 13(3), pp. 307–318. doi: 10.1007/s00500-008-0323-y.
- Algehyne, E. A. *et al.* (2022). Fuzzy neural network expert system with an improved gini index random forest-based feature importance measure algorithm for early diagnosis of breast cancer in saudi arabia. *Big Data and Cognitive Computing*, 6(13), 1-6. <https://doi.org/10.3390/bdcc6010013>
- Altay, A. dan Cinar, D. Fuzzy Decision Trees. In *Fuzzy Statistical Decision-Making*. Springer International Publishing, Switzerland, 2016. doi: 10.1007/978-3-319-39014-7.
- Alzoman, R. M. dan Alenazi, M. J. F. (2021). A comparative study of traffic classification techniques for smart city networks. *Sensors*, 21(14), 1–17. doi: 10.3390/s21144677.
- Andriansyah, M. R., Santoso, E. dan Sutrisno (2018). Klasifikasi risiko hipertensi menggunakan fuzzy decision tree iterative dichotomiser 3 (id3). *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 2(12), 7088–7096.
- Antini, C. *et al.* (2024). Diabetes mortality: trends and multi-country analysis of the the americas from 2000 to 2019. *Journal of Epidemiology*, 00,1–9. <https://doi.org/10.10 Thesis>
- Araniba, L. A. Q. (1994) *Learning Fuzzy Logic From Examples*. Thesis, Ohio University, USA.
- Barendregt, J. J. M. (1998). *Degenerative Disease in an Aging Population Models and Conjectures*. Fedobruk, Rotterdam, Belanda.
- Batista, P. and Pereira, A. (2016). Quality of life in patients with neurodegenerative diseases. *Journal of Neurology and Neuroscience*, 7(1), 1-7. doi: 10.21767/2171-6625.100074.
- Benarbia, M. (2022). A Machine Learning Approach to Predicting the Onset of Type II Diabetes in a Sample of Pima Indian Women Diabetes. Thesis City University of new York, USA.
- Bhattacharyya, R. and Mukherjee, S. (2020). Fuzzy membership function

- evaluation by non-linear regression: an algorithmic approach. *Fuzzy Information and Engineering*, 12(4), 412–434. <https://doi.org/10.1080/16168658.2021.1911567>
- Burger, S. V. (2018). *Introduction to Machine Learning with R: Rigorous Mathematical Analysis*. O'Reilly Media, Inc. USA.
- Çalışır, D. and Doğantekin, E. (2011). An automatic diabetes diagnosis system based on lda-wavelet support vector machine classifier. *Expert Systems with Applications*, 38(7), 8311–8315. <https://doi.org/10.1016/j.eswa.2011.01.017>.
- Carey, G. (1998). Multivariate analysis of variance (manova) ii : practical guide to anova and manova for sas. *Analysis*, pp. 1–14.
- Cavan, D., Makaroff, L. dan Fernandes, J. D. D. (2016). *Cost-Effective Solutions for The Prevention of Type 2 Diabetes*. Report of International Diabetes Federation.
- Chandrasekhar, N. dan Peddakrishna, S. (2023). Enhancing heart disease prediction accuracy through machine learning techniques and optimization. *Processes*, 11(1210), 1–31. <https://doi.org/10.3390/pr11041210>
- Chen, H. dan Fu, D. (2018). An Improved Naïve Bayes Classifier for Large Scale Text. *Proceedings of the 2018 2nd International Conference on Artificial Intelligence: Technologies and Applications (ICAITA)*, 2018.
- Chowdary, G. J. (2020). Effective Prediction of Cardiovascular Disease using Cluster of Machine Learning Algorithms. *Journal of Critical Reviews*, 7(19), 1865–1875.
- Conover, W. J., Johnson, M. E. dan Johnson, M. M. (1981). A comparative study of tests for homogeneity of variances, with applications to the outer continental shelf bidding data. *Technometrics*, 23(4), 351–361. doi: 10.2307/1268225.
- David, H. B. F. dan Belcy, S. A. (2018). Heart disease prediction using data mining techniques. *Ictact Journal on Soft Computing*, 09(01), 1817–1823. doi: 10.21917/ijsc.2018.0253.
- Di Cesare, M. *et al.* (2023). *World Heart Report 2023*. World Heart Federation: Geneva, Switzerland.
- Dougherty, J., Kohavi, R. dan Sahami, M. Supervised and Unsupervised Discretization of Continuous Features. *Proceedings of the Twelfth International Conference*, 1995.
- Eadie, M. J. (1974). Degenerative disease affecting the nervous system. *Australian Journal of Physiotherapy*, 20(1), 20–22. doi: 10.1016/S0004-9514(14)61167-3.
- Femina, B. T. dan Sudheep, E. M. (2020). A novel fuzzy linguistic fusion approach to naive bayes classifier for decision making applications. *International Journal on Advanced Science Engineering Information Technology*, 10(5), 1889–1897.
- Fernandez, S., Ito, T. dan Cruz-piris, L. (2022). Fuzzy ontology-based system for driver behavior classification. *Sensors*, 22(7954), 1–19. <https://doi.org/10.3390/s22207954>
- García, S., Luengo, J. dan Herrera, F. (2015). *Data Preprocessing in Data Mining*. Springer: Switzerland.

- Ghatak, A. (2017). *Machine Learning with R*. Springer Nature: Singapore; pp. 1–210. doi: 10.1007/978-981-10-6808-9.
- Han, J., Kamber, M. dan Pei, J. (2014). *Data Mining Concepts and Techniques*. Edisi Ketiga. Morgan Kaufmann: MA USA; pp. 203-207. doi: 10.1109/ICMIRA.2013.45.
- Harahap, J., Andayani, L. S. dan Utara, S. (2015). Screening of Degenerative Diseases and Quality of Life among Elderly People in Posyandu Lansia Medan Amplas. Proceedings of The 5th Annual International Conference Syiah Kuala University (AIC Unsyiah) 2015 In conjunction with The 8th International Conference of Chemical Engineering on Science and Applications (ChESA), 9-11 September 2015, Banda Aceh, Indonesia, pp. 295–299.
- Hassan, C. A. ul *et al.* (2022). Effectively predicting the presence of coronary heart disease using machine learning. *Sensors*, 22(19), 1-19. doi: 10.3390/s22197227.
- Hastie, T., Tibshirani, R. dan Friedman, J. (2017) *The Elements of Statistical Learning*. Springer: Switzerland; 83-85.
- Hossen, M. K. (2022). Heart disease prediction using machine learning techniques. *American Journal of Computer Science and Technology*, 5(3), 146-154. doi: 10.11648/j.ajcst.20220503.11
- IDF. (2017). *Diabetes Voice*. International Diabetes Federation: Brussels, Belgium.
- James, G. *et al.* (2013) *An Introduction to Statistical Learning with Application in R*. Springer: New York, USA.
- Kaggle. (2020). Cleveland Clinic Heart Disease Dataset. Diakses tanggal 8 Februari 2023 pada <https://www.kaggle.com/datasets/aavigan/cleveland-clinic-heart-disease-dataset/data>.
- Kaggle. (2016). Pima Indians Diabetes Datasets. Diakses tanggal 8 Februari 2023 pada <https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database>
- Kanwal, A. *et al.* (2022). Detection of heart disease using supervised machine learning. *VFAST Transactions on Software Engineering*, 6246, 58–70.
- Karthikeyini, S. *et al.* (2023). Heart disease prognosis using d-gru with logistic chaos honey badger optimization in iomt framework. *Information Technology and Control*, 52(2), 367–380. doi: 10.5755/j01.itc.52.2.32899.
- Kee, O. T. *et al.* (2023). Cardiovascular complications in a diabetes prediction model using machine learning: a systematic review. *Cardiovascular Diabetology*, 52(2), 1–10. doi: 10.1186/s12933-023-01741-7.
- Khamidah, F. S. N., Hapsari, D. P. dan Nugroho, H. (2018). Implementasi fuzzy decision tree untuk prediksi gagal ginjal kronis. *INTEGER: Journal of Information Technology*, 3(1), 19–28. doi: 10.31284/j.integer.2018.v3i1.155.
- Kresnawati, E. S. *et al.* (2021). Coronary artery disease prediction using decision trees and multinomial naïve bayes with k-fold cross validation. *Inomatika*, 3(2), 174–189. doi: 10.35438/inomatika.v3i2.266.
- Kresnawati, E. S., Suprihatin, B. and Resti, Y. (2024). The combinations of fuzzy membership functions on discretization in the decision tree-id3 to predict degenerative disease status. *Symmetry*, 16(1560), 1-38.

- <https://doi.org/10.3390/sym16121560>
- Kresnawati et al (2023). Diabetes Mellitus Diagnosis Using The Prediction Model of Discriminant Analysis. AIP Conference Proceedings of Annual Conference on Science and Technology Research, 2021, Palembang, Indonesia.
- Kumar, V. dan Chadha, A. (2012). Mining association rules in students assessment data. *International Journal of Computer Science Issues*, 9(5), 211–216.
- Kuo, N. I. *et al.* (2023). Synthetic health-related longitudinal data with mixed-type variables generated using diffusion models. *NeurIPS*, 1–24.
- Lantz, B. (2013). *Machine Learning with R, Machine Learning with R*. PACKT Publishing: Birmingham, UK. doi: 10.1002/9781119183464.ch12.
- Le, K. T. *et al.* (2020). An adapted linear discriminant analysis with variable selection for the classification in high-dimension, and an application to medical data. *Computational Statistics & Data Analysis*, 152, 1-18. doi: <https://doi.org/10.1016/j.csda.2020.107031>.
- Lê, S. (2019). Applied Multivariate Statistics. Dalam *Sensory Evaluation of Sound*. CRC Press: USA. doi: 10.1201/9780429429422-7.
- Lee, C. F., Tzeng, G. H. dan Wang, S. Y. (2005). A new application of fuzzy set theory to the Black-Scholes option pricing model. *Expert Systems with Applications*, 29(2), 330–342. doi: 10.1016/j.eswa.2005.04.006.
- Levashenko, V. and Zaitseva, E. (2012). Fuzzy Decision Trees in Medical Decision Making Support System. Prosiding Federated Conference on Computer Science and Information Systems, FedCSIS 2012, pp. 213–219. doi: 10.1177/2327857919081009.
- Livingston, K. A. *et al.* (2021). Lifestyle medicine and economics : a proposal for research priorities informed by a case series of disease reversal. *International Journal of Environmental Research and Public Health*, 18(11364), 1-13. <https://doi.org/10.3390/ijerph182111364>
- Maniruzzaman, M. *et al.* (2017). Comparative approaches for classification of diabetes mellitus data: Machine learning paradigm. *Computer Methods and Programs in Biomedicine*, 152, 33–34. doi: <https://doi.org/10.1016/j.cmpb.2017.09.004>.
- Mayfield, J. (1998). Diagnosis and classification of diabetes mellitus: new criteria. *American Family Physician*, 58(6), 1355-1362.
- Medasani, S. dan Kim, J. (1998). An overview of membership function generation techniques for pattern recognition. *International Journal of Approximate Reasoning*, 19, 391–417.
- Miltiadous, A. *et al.* (2023). Machine learning algorithms for epilepsy detection based on published eeg databases: a systematic review. *IEEE Access*, 11, 564–594. doi: 10.1109/ACCESS.2022.3232563.
- Muljadi, M., Suhaedi, D. dan Ramdani, Y. (2019). Analisis Konstruksi Fuzzy Decision Tree dengan Menggunakan Algoritma Iterative Dichotomiser 3 pada Aturan Klasifikasi Data Status Gizi. Prosiding Matematika, 2019.
- Olaru, C. and Wehenkel, L. (2003). A complete fuzzy decision tree. *Fuzzy Sets and Systems*, 138(2), 221-54.
- Retnasari, K. dan Rahmawati, E. (2017). Diagnosa Prediksi Penyakit Jantung

- Dengan Model Algoritma Naïve Bayes Dan Algoritma C4.5. Prosiding Konferensi Nasional Ilmu Sosial dan Teknologi (KNiST), 2017.
- Nelwan, E. J. *et al.* (2016). Modified risk factors for coronary heart disease (chd) in minahasa ethnic group from manado city indonesia. *Journal of Experimental Life Science*, 6(2), 88–94.
- Nezhad, S. N., Zahedi, M. H. dan Farahani, E. (2022). Detecting diseases in medical prescriptions using data mining methods. *BioData Mining*, 15(29), 1–20. doi: 10.1186/s13040-022-00314-w.
- Palanivinayagam, A. dan Damaševičius, R. (2023). Effective handling of missing values in datasets for classification using machine learning methods. *Information*, 14(2), 1–15. doi: 10.3390/info14020092.
- Palatini, P. (1999). Need for a revision of the normal limits of resting heart rate. *Hypertension*, 33(2), 622-625. doi: 10.1161/01.hyp.33.2.622.
- Patil, S. dan Bhosale, S. (2023). Improving cardiovascular disease prognosis using outlier detection and hyperparameter optimization of machine learning models. *Revue d ' Intelligence Artificielle*, 37(4), 1069–1080.
- Purushottam, Saxena, K. and Sharma, R. (2016). Efficient heart disease prediction system. *Procedia Computer Science*, 85, 962–969. doi: 10.1016/j.procs.2016.05.288.
- Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, 1(1), 81–106. doi: 10.1007/bf00116251.
- Raghavendra, U. *et al.* (2016). application of gabor wavelet and locality sensitive discriminant analysis for automated identification of breast cancer using digitized mammogram images. *Applied Soft Computing Journal*, 46(May), 151–161. doi: 10.1016/j.asoc.2016.04.036.
- Ramasubramanian, K. and Singh, A. (2019) *Machine Learning Using R*. Apress: California, USA.
- Renzo, L. Di *et al.* (2023). Nutrition for prevention and control of chronic degenerative diseases. *Nutrients*, 15(2253), 8–11. doi: 10.3390/nu15102253
- Resti, Y. *et al.* (2020). Improved the cans waste classification rate of naive bayes using fuzzy approach. *Science and Technology Indonesia*, 5(3), 75–78.
- Resti, Y. *et al.* (2021). Diagnosis of diabetes mellitus in women of reproductive age using the prediction methods of naive bayes, discriminant analysis, and logistic regression. *Science and Technology Indonesia*, 6(2), 96–104. doi: 10.26554/STI.2021.6.2.96-104.
- Resti, Y., *et al.* (2022). Identification of corn plant diseases and pests based on digital images using multinomial naïve bayes and k-nearest neighbor. *Science and Technology Indonesia*, 7(1), 29–35. doi: 10.26554/sti.2022.7.1.29-35.
- Resti, Y., *et al.* (2022). performance improvement of decision tree model using fuzzy membership function for classification of corn plant diseases and pests. *Science and Technology Indonesia*, 7(3), 284-290. <https://doi.org/10.26554/sti.2022.7.3.284-290>
- Resti, Y. *et al.* (2023). Fuzzy Discretization on the Multinomial Naïve Bayes Method for Modeling Multiclass Classification of Corn Plant Diseases and Pests. *Mathematics*, 11(1761), 1-21. <https://doi.org/10.3390/math11081761>

- Resti, Y. (2024). Credit Risk-Type Classification using Statistical Learning. Prosiding dari the 3rd Conference on Fundamental and Applied Science for Advanced Technology 2022, Palembang, Indonesia.
- Riani, A., Susianto, Y. dan Rahman, N. (2019). Implementasi data mining untuk memprediksi penyakit jantung menggunakan metode naive bayes. *Journal of Innovation Information Technology and Application (JINITA)*, 1(01), 25–34. doi: 10.35970/jinita.v1i01.64.
- Snedecor, G. dan Cochran, W. (1989). *Statistical Methods*, Edisi Kedelapan. Iowa State University Press: Iowa, USA.
- Rodríguez, J. D., Martínez, A. P. and Lozano, J. A. (2010). Sensitivity analysis of k-fold cross validation in prediction error estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(3), 569–575.
- Roy, A. and Pal, S. K. (2003). Fuzzy discretization of feature space for a rough set classifier. *Pattern Recognition Letters*, 24, 895–902 24.
- Rutkowski, L. (2004). *Flexible Neuro-Fuzzy Systems*. Kluwer Academic Publisher: Boston, USA.
- Santoso, H. (2012). Analisis dan Prediksi Perilaku Mahasiswa Diploma untuk Melanjutkan Studi ke Jenjang Sarjana menggunakan Teknik Decision Tree dan Support Vector Machine. Thesis, Universitas Sumatera Utara, Indonesia.
- Selvarathi, C. dan Varadhaganapathy, S. (2023). Deep learning based cardiovascular disease risk factor prediction among type 2 diabetes mellitus patients. *Information Technology and Control*, 52(1), 215–227. doi: 10.5755/j01.itc.52.1.32008.
- Shanmugapriya, M. *et al.* (2017). Fuzzy discretization based classification of medical data. *Research Journal of Applied Sciences, Engineering and Technology*, 14(8), 291-298. doi: 10.19026/rjaset.14.4953.
- Székely, G. J. dan Rizzo, M. L. (2005). A new test for multivariate normality. *Journal of Multivariate Analysis*, 93(1), 58–80. doi: <https://doi.org/10.1016/j.jmva.2003.12.002>.
- Tasin, I. *et al.* (2023). Diabetes prediction using machine learning and explainable AI. *Healthcare Technology Letters*, 10, 1–10. doi: 10.1049/htl2.12039.
- Tigga, N. P. dan Garg, S. (2020). Prediction of type 2 diabetes using machine learning classification methods. *Procedia Computer Science*, 167, 706–716.
- Tutuncu, G. Y. dan Kayaalp, N. (2019). An aggregated fuzzy naive bayes data classifier. *Journal of Computational and Applied Mathematics*, 286, 17–27.
- Wang, L. *et al.* (2016). The Use of Fuzzy Logic Techniques to Improve Decision Making in Apparel Supply Chains. *Information Systems for the Fashion and Apparel Industry*. Woodhead Publishing: MA, USA, pp. 9–39. doi: 10.1016/B978-0-08-100571-2.00002-6.
- WHO. (2016). *Global Report on Diabetes*. World Health Organization: Geneva, Switzerland.
- WHO (2019). *World Health Statistics Overview 2019*. World Health Organization: Geneva, Switzerland.
- Witten, I. H. dan Frank, E. (2005). *Data Mining Practical Machine Learning Tools and Techniques*. Edisi Kedua. Morgan Kaufmann Publisher: California, CA, USA.

