

**MEMPREDIKSI RISIKO PENYAKIT JANTUNG MENGGUNAKAN
METODE *DECISION TREE* DENGAN VALIDASI SILANG**

SKRIPSI

**Sebagai Salah Satu Syarat untuk Memperoleh Gelar
Sarjana Matematika Bidang Studi Matematika**



Oleh :

**MUHAMMAD RENDY KURNIAWAN
NIM 08011281621074**

**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS SRIWIJAYA
JANUARI 2021**

LEMBAR PENGESAHAN

**MEMPREDIKSI RISIKO PENYAKIT JANTUNG MENGGUNAKAN
METODE *DECISION TREE* DENGAN VALIDASI SILANG**

SKRIPSI

**Sebagai Salah Satu Syarat untuk Memperoleh Gelar
Sarjana Matematika Bidang Studi Matematika**

Oleh

**MUHAMMAD RENDY KURNIAWAN
NIM 08011281621074**

Indralaya, Januari 2021

Pembimbing Kedua



**Endang Sri Kresnawati, M.Si
NIP. 197702082002122003**

Pembimbing Utama



**Dr. Yulia Resti, M.Si
NIP. 197307191997022001**

Mengetahui

Ketua Jurusan Matematika



**Dr. Sugandi Yahdin, M.M
NIP. 195807271986031003**

HALAMAN PERSEMBAHAN

MOTTO

"..Karena sesungguhnya sesudah kesulitan itu ada kemudahan. Sesungguhnya sesudah kesulitan itu ada kemudahan.." (Qs. Al-Insyirah (94) : 5-6)

“Barangsiapa melepaskan kesusahan duniawi seorang Muslim, Allah akan melepaskan kesusahannya pada hari kiamat. Barangsiapa memudahkan seorang yang mendapat kesusahan, Allah akan memudahkan urusannya di dunia dan akhirat.” (HR Muslim)

Dua rakaat (sebelum) shalat fajar (subuh) lebih baik dari dunia dan seisinya.
(HR. Muslim).

“Hargailah waktu, karena waktu sangat berharga”

Skripsi ini kupersembahkan untuk :

- ✓ Allah SWT
- ✓ Ayah, Ibu, dan Saudara-saudaraku
- ✓ Dosenku
- ✓ Keluarga besarku
- ✓ Sahabat dan Teman-temanku
- ✓ Kampusku

KATA PENGANTAR

Assalamu'alaikum warahmatullahi wabarakatuh

Puji syukur kehadirat Allah SWT Yang Maha Pengasih lagi Maha Penyayang atas limpahan rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul “**Memprediksi Risiko Penyakit Jantung Menggunakan Metode *Decision Tree* dengan Validasi Silang**” tepat pada waktunya. Shalawat serta salam semoga senantiasa tercurahkan kepada Nabi Muhammad SAW beserta keluarga, para sahabat, serta seluruh pengikutnya hingga akhir zaman. Skripsi ini sebagai salah satu syarat untuk memperoleh gelar Sarjana Sains bidang studi Matematika di Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Sriwijaya.

Dengan segala hormat dan kerendahan hati, penulis mengucapkan terima kasih yang tak terhingga kepada orang tua tercinta, yaitu **Ibu Bar'kah** dan **Bapak Rizik** yang telah merawat, menasehati dengan penuh rasa cinta dan kasih sayang, serta dukungan yang sangat berharga berupa doa, motivasi, perhatian, serta material untuk penulis selama ini. Penulisan skripsi ini tidak terlepas bantuan berbagai pihak, baik secara langsung maupun tidak langsung. Untuk itu penulis menyampaikan ucapan terima kasih dan penghargaan kepada :

1. **Kemenristekdikti** selaku pemberi bantuan beasiswa Bidikmisi selama penulis menempuh pendidikan di Universitas Sriwijaya dari awal perkuliahan hingga penulis dapat menyelesaikan jenjang Strata Satu (S1) Universitas Sriwijaya.

2. Bapak **Drs. Sugandi Yahdin, M.M.**, selaku Ketua Jurusan Matematika atas bimbingan dan ilmu yang bermanfaat yang telah diberikan selama penulis belajar di Jurusan Matematika.
3. Ibu **Des Alwine Zayanti, M.Si.**, selaku Sekretaris Jurusan Matematika dan selaku Dosen Pembahas yang telah bersedia meluangkan waktu dalam memberikan tanggapan, kritik dan saran yang bermanfaat dalam perbaikan dan penyelesaian skripsi ini.
4. Ibu **Dr. Yulia Resti, M.Si.**, selaku Dosen Pembimbing Utama yang telah bersedia meluangkan waktu, pikiran, tenaga, dan nasehat dalam membimbing serta mengarahkan penulis dalam penyelesaian skripsi sehingga penulis dapat menyelesaikan skripsi ini.
5. Ibu **Endang Sri Kresnawati, M.Si.**, selaku Dosen Pembimbing Kedua dan Dosen Pembimbing Akademik yang telah bersedia meluangkan waktu, pikiran, tenaga, dan nasehat dalam membimbing serta mengarahkan penulis dalam akademik dan penyelesaian skripsi sehingga penulis dapat menyelesaikan skripsi ini.
6. Bapak **Dr. Bambang Suprihatin, M.Si.**, dan Ibu **Oki Dwipuwani, M.Si.**, selaku Dosen Pembahas yang telah bersedia meluangkan waktu dalam memberikan tanggapan, kritik dan saran yang bermanfaat dalam perbaikan dan penyelesaian skripsi ini.
7. **Seluruh Dosen di Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Sriwijaya** yang telah memberikan ilmu

yang bermanfaat, nasehat, serta bimbingan selama penulis menjalani perkuliahan.

8. Bapak **Irwansyah** selaku admin dan Ibu **Khamidah** selaku pegawai tata usaha Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Sriwijaya yang telah membantu penulis selama masa perkuliahan.
9. Orang tuaku tercinta Bapak **A.Rizik.A.Halim** dan Ibu **Bar'kah Hairulnisya**, Kakakku dan Adikku tersayang **Muhammad Reza** dan **Muhammad Renaldy Irfani** serta keluargaku yang selalu memberikan kasih sayang, doa, motivasi, bantuan, dan nasihatnya.
10. Sahabat sepermabaan **Akmal, Miko, Karita, Widya, Tissa, dan Nopita**, Sahabat Skripsweet Squad **Megawati, Tiya, Widya Ayu, Indah, Hariani, Eling, Nana, Ari, Jekta** dan **Eko**, Teman se-PA **Danial** dan **Aka**, Teman akrabku **Naurah, Anita, Shania, Neysa, Septia, Fero, Jihan, Priska, Anbil, Dea, Delia, Giska, Irvan, Fathur, Anton, Doni, Ogi** dan **Rahmat**, serta teman-teman seperjuangan Angkatan 2016 lainnya untuk semua bantuan, canda tawa, suka duka, ketulusan, motivasi, pikiran, nasehat dan doanya.
11. Kakak tingkat 2015, 2014, 2013, 2012 yang telah memberikan arahan dan bantuan selama perkuliahan serta dalam penulisan skripsi ini.
12. Adik-adik tingkatku 2017 **Indri, Sari, Abu, Kahfi, Fauzi, Yudha, Juli** dan lainnya, Keluarga Bergema dan Departemen Sosmas 2018-2019. Adik-adik asuhku “Keluarga Dilan” **Indrike** dan **Chatrin** serta adik-adik angkatan

2018 lainnya yang telah memberikan dukungan dan semangat selama perkuliahan

13. Sahabat 3 Serangkaiku **Refzqi Safei** dan **Keken Wilandra** dan Sahabat Hyna **Nang Yudhi, Gudel, Hendri, Idrus, Honest, Wowok** dan **Angwin** serta Sahabat 3K yaitu **Fahmi** dan **Umar** untuk semua bantuan, semangat dan motivasi.
14. Sahabat KKN 92 Kota Raya “Cecingal Squad” **Syahrul, Rudi, Widya Ayu, Penty, Vaya, Diana, Viona, Jija, Tanty** dan **Icak** atas 40 hari mengabdikan bersama serta memberikan semangat dalam penulisan skripsi ini.
15. Sahib-sahibku **Kak Ren, Bang Rul, Mas Rud,** dan **A’Den** telah menjadi teman yang membawa dan mengajak kepada kebaikan.
16. **Semua pihak** yang telah membantu penulis dalam menyelesaikan skripsi ini yang tidak dapat disebutkan satu per satu, sehingga segala kebaikannya mendapat pahala dari Allah SWT.

Semoga skripsi ini dapat bermanfaat bagi yang membacanya.

Wassalamu 'alaikum warahmatullahi wabarakatuh

Indralaya, Januari 2021

Penulis

PREDICTING THE RISK OF HEART DISEASE USING THE DECISION TREE METHOD WITH CROSS VALIDATION

By:

Muhammad Rendy Kurniawan
08011281621074

ABSTRACT

According to data from the World Health Organization (WHO) in 2015, more than 17 million people worldwide died from heart and blood vessel disease, which is the highest death rate in the world. Heart disease occurs due to narrowing and obstruction of blood vessels so that the supply of oxygen and nutrients to the heart is reduced. If symptoms of heart disease are known, anticipation can be done. Predicting the risk of heart disease aims to see whether a person has heart disease or not from the symptoms it causes. In this study, the decision tree method was used with cross validation to predict the risk of heart disease based on 13 variables that affect heart disease, namely age, sex, chest pain, trestbps, chol, fasting blood sugar, restecg, thalach, exang, oldpeak, slope, ca, thal. Based on the calculation results, the variables that affect the risk of heart disease are chest pain, ca and thal and the average overall classification accuracy rate with cross validation of 75.77%.

Keywords: Heart Disease, Decision Tree, C4.5 Algorithm

Indralaya, Januari 2021

Pembimbing Kedua



Endang Sri Kresnawati, M.Si
NIP. 197702082002122003

Pembimbing Utama



Dr. Yulia Resti, M.Si
NIP. 197307191997022001

MEMPREDIKSI RISIKO PENYAKIT JANTUNG MENGGUNAKAN METODE DECISION TREE DENGAN VALIDASI SILANG

Oleh

Muhammad Rendy Kurniawan
08011281621074

ABSTRAK

Menurut data Organisasi Kesehatan Dunia (WHO) tahun 2015, lebih dari 17 juta orang di dunia meninggal akibat penyakit jantung dan pembuluh darah yang merupakan penyebab kematian tertinggi di dunia. Terjadinya penyakit jantung karena penyempitan dan penghambatan pembuluh darah jantung sehingga penyaluran oksigen dan nutrisi ke jantung berkurang. Jika gejala penyakit jantung diketahui maka dapat dilakukan antisipasi. Memprediksi risiko penyakit jantung bertujuan untuk mengetahui seseorang terkena penyakit jantung atau tidak dari gejala yang ditimbulkan. Pada penelitian ini digunakan metode *decision tree* dengan validasi silang untuk memprediksi risiko penyakit jantung berdasarkan 13 variabel yang mempengaruhi penyakit jantung yaitu umur, jenis kelamin, sakit dada, tekanan darah tinggi (hipertensi), kadar kolestrol, kadar glukosa (gula darah), hasil EKG selama istirahat, detak jantung maksimal yang dicapai, ukuran *boolean*, segmen ST, kemiringan segmen ST, jumlah vessel utama diwarnai oleh fluoroskopi, dan status jantung. Berdasarkan hasil perhitungan diperoleh variabel yang berpengaruh terhadap risiko penyakit jantung adalah sakit dada, jumlah vessel utama diwarnai oleh fluoroskopi, dan status jantung serta rata-rata tingkat akurasi klasifikasi secara keseluruhan dengan validasi silang sebesar 75.77%.

Kata kunci : Penyakit Jantung, *Decision Tree*, Algoritma C4.5

Indralaya, Januari 2021

Pembimbing Kedua



Endang Sri Kresnawati, M.Si
NIP. 197702082002122003

Pembimbing Utama



Dr. Yulia Resti, M.Si
NIP. 197307191997022001

DAFTAR ISI

	Halaman
HALAMAN JUDUL	i
HALAMAN PENGESAHAN	ii
HALAMAN PERSEMBAHAN	iii
KATA PENGANTAR	iv
ABSTRACT	viii
ABSTRAK	ix
DAFTAR ISI	x
DAFTAR TABEL	xiii
DAFTAR GAMBAR	xiv
DAFTAR LAMPIRAN	xv
BAB I PENDAHULUAN	
1.1 Latar Belakang	1
1.2 Perumusan Masalah	5
1.3 Pembatasan Masalah	5
1.4 Tujuan	5
1.5 Manfaat	6
BAB II TINJAUAN PUSTAKA	
2.1 <i>Data Mining</i>	7
2.1.1 Pembelajaran Mesin (<i>Machine Learning</i>)	7
2.1.2 Pengelompokan <i>Data Mining</i>	7
2.2 Klasifikasi	8

2.3	<i>Decision Tree</i>	10
2.3.1	Model <i>Decision Tree</i>	10
2.3.1	Algoritma C4.5.....	11
2.4	Pengujian dan Evaluasi Model.....	14
2.5	Penyakit Jantung	15
2.5.1	Gejala Penyakit jantung	16
2.5.2	Faktor Risiko Penyakit Jantung	17
2.5.3	Diagnosa Penyakit.....	21

BAB III METODOLOGI PENELITIAN

3.1	Tempat.....	28
3.2	Waktu	28
3.3	Metode Penelitian.....	28

BAB IV HASIL DAN PEMBAHASAN

4.1	Deskripsi Data.....	30
4.2	Diskritisasi Data	32
4.3	<i>Fold 1</i>	33
4.4	Hasil untuk <i>Fold 2</i> sampai <i>Fold 5</i>	46
4.4.1	<i>Fold 2</i>	47
4.4.2	<i>Fold 3</i>	47
4.4.3	<i>Fold 4</i>	48
4.4.4	<i>Fold 5</i>	49
4.5	Pengujian dan Evaluasi Model.....	50

BAB V KESIMPULAN DAN SARAN

5.1 Kesimpulan	53
5.2 Saran.....	53
DAFTAR PUSTAKA	54
LAMPIRAN.....	57

DAFTAR TABEL

	Halaman
Tabel 2.1 <i>Confusion Matrix</i> 2 kelas	15
Tabel 2.2 Klasifikasi Usia	23
Tabel 2.3 Klasifikasi <i>Chest Pain</i>	23
Tabel 2.4 Klasifikasi <i>Trestbps</i>	24
Tabel 2.5 Klasifikasi Kolestrol	24
Tabel 2.6 Klasifikasi FBS	25
Tabel 2.7 Klasifikasi <i>Restecg</i>	25
Tabel 2.8 Klasifikasi <i>Thalach</i>	25
Tabel 2.9 Klasifikasi <i>Exang</i>	26
Tabel 2.10 Klasifikasi <i>Oldpeak</i>	26
Tabel 2.11 Klasifikasi <i>Slope</i>	26
Tabel 2.12 Klasifikasi <i>Thal</i>	27
Tabel 4.1 Ringkasan Data Variabel Numerik	31
Tabel 4.2 Ringkasan Data Variabel Kategori	31
Tabel 4.3 Diskritisasi Data	32
Tabel 4.4 <i>Confusion Matrix Testing Fold 1 dan Training Fold 2</i>	51
Tabel 4.5 Rata-rata Tingkat Akurasi dari Proses Validasi Silang untuk <i>Testing Fold 1</i>	51
Tabel 4.6 Rata-rata Tingkat Akurasi Secara Keseluruhan dari Proses Validasi Silang	51

DAFTAR GAMBAR

	Halaman
Gambar 2.1 Konsep <i>Decision Tree</i>	10
Gambar 2.2 Model <i>Decision Tree</i>	11
Gambar 2.3 Ilustrasi <i>4-Fold Cross Validation</i>	14
Gambar 4.1 Pohon Keputusan <i>Node Akar</i>	36
Gambar 4.2 Pohon Keputusan <i>Node Cabang 1.1</i>	37
Gambar 4.3 Pohon Keputusan <i>Node Cabang 1.3</i>	38
Gambar 4.4 Pohon Keputusan <i>Node Cabang 1.3.1</i>	39
Gambar 4.5 Pohon Keputusan <i>Node Cabang 1.3.1.2</i>	40
Gambar 4.6 Pohon Keputusan <i>Fold 1</i>	43

DAFTAR LAMPIRAN

	Halaman
Lampiran 1 Pohon Keputusan <i>Fold 2</i>	57
Lampiran 2 Pohon Keputusan <i>Fold 3</i>	58
Lampiran 3 Pohon Keputusan <i>Fold 4</i>	59
Lampiran 4 Pohon Keputusan <i>Fold 5</i>	60

BAB I

PENDAHULUAN

1.1 Latar Belakang

Jantung adalah salah satu organ tubuh yang berperan sangat penting bagi manusia karena berfungsi memompa darah ke seluruh tubuh (Aulia, 2018). Menurut data Organisasi Kesehatan Dunia (WHO) tahun 2015 bahwa lebih dari 17 juta orang di dunia meninggal akibat penyakit jantung dan pembuluh darah yang merupakan penyebab kematian tertinggi di dunia.

Terjadinya penyakit jantung karena penyempitan dan penghambatan pembuluh darah jantung sehingga penyaluran oksigen dan nutrisi ke jantung berkurang. Jika gejala penyakit jantung diketahui maka dapat dilakukan antisipasi. Dengan memprediksi risiko penyakit jantung dapat mengetahui seseorang terkena penyakit jantung atau tidak dari gejala yang ditimbulkan.

Menurut Larose (2005) bahwa *data mining* adalah suatu proses menemukan hubungan yang berarti, pola, dan kecenderungan dengan adanya pemeriksaan dari sekumpulan besar data yang tersimpan dengan menggunakan teknik pengenalan pola yaitu teknik statistik dan matematika. Ada suatu metode di dalam *data mining* yang disebut *Classification*, yaitu metode untuk memprediksi atau pengklasifikasian kategori dari suatu data berdasarkan sekumpulan variabel atau atribut dari data tersebut.

Salah satu metode klasifikasi yang paling populer dan mudah untuk dipahami oleh manusia adalah *decision tree*. Heri Santoso (2012), menjelaskan

bahwa *decision tree* merupakan salah satu metode klasifikasi yang menggunakan representasi struktur pohon (*tree*) di mana setiap *node* menjelaskan variabel, cabangnya menjelaskan nilai dari variabel, dan daun menjelaskan kelas. *Decision tree* mempunyai tingkat akurasi yang cukup tinggi pada berbagai kasus.

Banyak algoritma yang digunakan untuk pembentukan pohon keputusan antara lain ID3, CART, dan C4.5. Algoritma C4.5 merupakan pengembangan dari algoritma ID3 (Larose, 2005).

Algoritma C4.5 adalah kelompok algoritma *decision tree*. Algoritma ini memiliki *input* seperti *training samples* dan *samples*. *Training samples* adalah data contoh yang akan digunakan untuk membangun sebuah pohon yang telah diuji kebenarannya. Sedangkan *samples* adalah variabel yang nantinya akan digunakan sebagai parameter dalam melakukan klasifikasi data (Sunjana, 2010)

Pada umumnya, algoritma yang digunakan untuk membentuk *decision tree* adalah algoritma C4.5. Kelebihan algoritma C4.5 sendiri yaitu dapat menghasilkan pohon keputusan dengan tingkat akurasi yang dapat diterima dan efisien dalam menangani variabel bertipe diskrit atau numerik (Pangaribuan dkk., 2019)

Pada tugas akhir ini penulis akan memprediksi risiko penyakit jantung menggunakan metode *decision tree* serta melakukan pengujian dan evaluasi model berdasarkan ketepatan data atau tingkat akurasi.

Menurut Normawati dan Winiarti (2017) bahwa dari hasil perbandingan proses pengujian diagnosis penyakit jantung koroner dengan metode seleksi fitur menggunakan penambahan data berbasis *Variable Precision Rough Set* (VPRS)

dengan 14 variabel dan 296 data pengamatan adalah 84,84%, sedangkan *Motivated Feature Selection* (MFS) adalah 86,86% dan MFS+VPRS adalah 84,84% dengan $\frac{2}{3}$ data latih dan $\frac{1}{3}$ data uji.

Menurut Retnasari dan Rahmawati (2017) bahwa Algoritma Naïve Bayes memiliki nilai akurasi yang paling tinggi yaitu 86.67% sedangkan algoritma C4.5 memiliki nilai akurasi 83.70% dengan 13 variabel kecuali usia dan 270 data pengamatan.

Menurut Aini dkk. (2018) bahwa hasil penelitian seleksi fitur *information gain* menggunakan kombinasi metode *K-Nearest Neighbor* dan *Naive Bayes* menunjukkan nilai akurasi tertinggi yang diperoleh saat menggunakan data latih dengan label kelas seimbang yaitu sebesar 92,31% dengan enam fitur dan nilai $K=25$. Sedangkan nilai akurasi tertinggi yang diperoleh saat menggunakan data latih dengan label kelas tidak seimbang yaitu sebesar 92,31% dengan empat fitur dan nilai $K=35$.

Menurut Indrajani dkk. (2018) bahwa pada diagnosa penyakit jantung dengan menggunakan *machine learning* dengan hasil prediksi penyakit menggunakan ROC, mendapat tingkat akurasi sebesar 82%, menggunakan model k-NN ($k = 7$).

Gavin Brown (2004) mengungkapkan tingkat kesalahan klasifikasi dari dua uji empiris pada data kanker payudara dan penyakit jantung menggunakan *Negative Correlation Learning* (NC) dengan *5-fold cross validation* dan 40 uji coba dari seragam bobot acak dalam $[-0.5, 0.5]$ untuk setiap pengaturan, total 200 uji coba dilakukan untuk masing-masing konfigurasi eksperimental. Perlu di catat

dengan 2 jaringan $\gamma = \lambda$ dan nilai λ yang diuji di NC : 0, 0.5, 1. Hasil pada data jantung menunjukkan NC secara signifikan lebih baik daripada ensemble sederhana (setara dengan $\lambda = 0$) dengan $\alpha = 0.05$ pada uji-t dua arah. Untuk data kanker payudara meskipun rata-rata lebih rendah, itu tidak signifikan secara statistik.

Menurut Chai dkk. (2004) bahwa dalam merancang pengklasifikasi yang meminimalkan jumlah biaya kesalahan klasifikasi dan biaya pengujian menggunakan *Test-Cost Sensitive Naive Bayes Classification* pada 8 dataset dari UCI ML repository didapatkan lebih memilih *csNaive Bayes* dalam *The Cost Sensitive* karena lebih unggul dalam meminimalkan biaya kesalahan klasifikasi dan biaya pengujian dibandingkan algoritma lainnya termasuk *csDecision Tree*.

Menurut Huang dkk. (2004) bahwa *The Minimax Probability Machine* (MPM) membuat pengklasifikasi, yang menyediakan batasan kasus terburuk pada probabilitas kesalahan klasifikasi dengan poin data di masa depan berdasarkan perkiraan yang dapat diandalkan dari matriks mean dan kovariansi kelas dari poin data *training*, dan mencapai kinerja komparatif dengan pengklasifikasi canggih, *The Support Vector Machine*. Dengan menghilangkan asumsi bobot tidak bias untuk setiap kelas di MPM dan mengembangkan ekstensi kritis, bernama *Biased Minimax Probability Machine* (BMPM), untuk ditangani tugas klasifikasi bias, terutama dalam aplikasi diagnostik medis. Kami menguraikan turunan teoritis dari BMPM. Selain itu, kami menunjukkan bahwa model ini dapat diubah menjadi pecahan cekung-cembung masalah pemrograman atau masalah pseudoconcave.

1.2 Rumusan Masalah

Berdasarkan uraian pada latar belakang yang telah dikemukakan sebelumnya, maka perumusan masalah pada penelitian ini adalah :

1. Bagaimana memprediksi risiko penyakit jantung menggunakan metode *decision tree* dengan validasi silang?
2. Berapa ketepatan data atau tingkat akurasi dalam memprediksi risiko penyakit jantung menggunakan metode *decision tree* dengan validasi silang?

1.3 Pembatasan Masalah

Pembatasan masalah dalam penelitian ini dibatasi oleh :

1. Data *Machine Learning Repository* dari *University of California Irvine* (UCI) Tahun 1988.
2. Data yang digunakan sebanyak 270 dari 303 pengamatan karena ada sebanyak 33 data pengamatan yang hilang artinya ada beberapa yang kosong dalam data tersebut sehingga tidak digunakan dalam tugas akhir ini.

1.4 Tujuan

Berdasarkan latar belakang yang telah dikemukakan, maka tujuan dari penelitian ini adalah :

1. Melakukan prediksi risiko penyakit jantung menggunakan metode *decision tree* dengan validasi silang.
2. Menghitung ketepatan data atau tingkat akurasi dalam memprediksi risiko penyakit jantung menggunakan metode *decision tree* dengan validasi silang.

1.5 Manfaat

Penelitian ini diharapkan dapat memberikan manfaat sebagai berikut :

1. Menambah pengetahuan penulis tentang memprediksi risiko penyakit jantung menggunakan metode *decision tree* dengan validasi silang.
2. Penelitian ini diharapkan dapat membantu dan mempermudah tenaga medis dalam memprediksi risiko penyakit jantung.
3. Sebagai referensi dan sumber pustaka untuk penulisan penelitian yang sejenis.

DAFTAR PUSTAKA

- Aini, S. H. A., Sari, Y. A., dan Arwan, Achmad. (2018). Seleksi Fitur Information Gain untuk Klasifikasi Penyakit Jantung Menggunakan Kombinasi Metode *K-Nearest Neighbor* dan *Naïve Bayes*. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer* Vol. 2, No. 9.
- Aulia, Wizra. 2018. Sistem Pakar Diagnosa Penyakit Jantung Koroner Dengan Metode *Probabilistic Fuzzy Decision Tree*. *Jurnal Sains dan Informatika*. 4(12):106-117.
- Brown, G. (2004). *Diversity in Neural Network Ensembles*. University of Birmingham, School of Computer Science.
- Buku Pintar POSBINDU PTM. 2016. *Penyakit Tidak Menular Dan Faktor Risiko*. Kemenkes RI.
- Chai, Xiaoyong & Deng, Lin & Yang, Qiang & Ling, C.X.. (2004). Test-Cost Sensitive Naive Bayes Classification. *Proceedings - Fourth IEEE International Conference on Data Mining, ICDM 2004*. 51- 58. 10.1109/ICDM.2004.10092.
- Emha Taufiq Luthfi dan Kusri, *Algoritma Data Mining*. Yogyakarta: Andi, 2009.
- Gorunescu, Florin. 2011. *Data Mining: concepts, models, and techniques* Verlag Berlin Heidelberg: Springer.
- Han, J., Kamber, M., dan Pei, J. (2012). *Data Mining: Concept and Techniques*, Third Edition. Waltham: Morgan Kaufmann.
- Han, J dan Kamber, M., “*Data Mining Concept and Technique*”, Morgan Kaufmann, 2001.
- Indrajani, dkk. “Aplikasi Informasi Kesehatan dan Diagnosa Penyakit Jantung Berbasis Android”, Seminar Nasional Teknologi Informasi, Komunikasi dan Industri (SNTIKI-10), November 2018.
- Jan Wira Gotama Putra. Pengenalan Konsep Pembelajaran Mesin dan Deep Learning. Self-published work, 2018. Diakses dari <https://wiragotama.github.io/resources/ebook/intro-to-ml-secured.pdf>, 28 April 2020.

- Kaizhu Huang, Haiqin Yang, Irwin King, Michael R. Lyu, and Laiwan Chan. Biased Minimax Probability Machine for Medical Diagnosis. The 8th International Symposium on Artificial Intelligence and Mathematics (AMAI'2004), Fort Lauderdale, Florida, January 4-6, 2004.
- Kholifah, Ahadiyah Nurul and Insani, Nur (2016). Analisis Klasifikasi Pada Nasabah Kredit Koperasi X Menggunakan *Decision Tree* C4.5 Dan *Naïve Bayes*. S1 thesis, UNY.
- Kurniadi, H, 2013. *Stop Gejala Penyakit Jantung Koroner*. Yogyakarta: Familia.
- Larose, Daniel T. 2005. *Discovering Knowledge in Data: An Introduction to Data Mining*. John Willey & Sons, Inc.
- Listiana, M. (2015). Perbandingan Algoritma *Decision Tree* (C4.5) dan *Naïve Bayes* Pada Data Mining Untuk Identifikasi Tumbuh Kembang Anak Balita (Studi Kasus Puskesmas Kartasura), Skripsi. Universitas Muhammadiyah Surakarta (pp. 4-5).
- Madadipouya, K. (2015). A New Decision Tree Method For Data Mining In Medicine. *Advanced Computational Intelligence An International Journal*, 2 (3), 33-35.
- Normawati, Dwi., dan Winiarti, Sri. (2017). Seleksi Fitur Menggunakan Penambangan Data Berbasis Variable Precision Rough Set (VPRS) Untuk Diagnosis Penyakit Jantung Koroner. *Jurnal Ilmu Teknik Elektro Komputer dan Informatika (JITEKI)* Vol. 3, No. 2
- Notoatmodjo S. Kesehatan Masyarakat : Ilmu dan Seni. Jakarta: PT Rineka Cipta; 2011.
- Pangaribuan J. J., Tedja, C., dan Wibowo, S. (2019). Perbandingan Metode Algoritma C4.5 Dan Extreme Learning Machine untuk Mendiagnosis Penyakit Jantung Koroner. *Informatics Engineering Research and Technology* Vol. 1, No.1
- Retnasari, Tri., dan Rahmaawati, Eva, "Diagnosa Prediksi Penyakit Jantung Dengan Model Algoritma *Naïve Bayes* dan Algoritma C4.5," Konferensi Nasional Ilmu Sosial & Teknologi (KNiST), pp. 7-12, Maret 2017.
- Santosa, B. 2007. *Data Mining: Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Graha Ilmu. Yogyakarta.

- Santoso, Heri. 2012. Analisis Dan Prediksi Pada Perilaku Mahasiswa Diploma Untuk Melanjutkan Studi Ke Jenjang Sarjana Menggunakan Teknik *Decision Tree* Dan *Support Vektor Machine* [Tesis]. Medan (ID) : Universitas Sumatera Utara.
- Suhartono. Analisis Data Statistik dengan R. Surabaya : Lab Statistik Komputasi; 2008.
- Sunjana, 2010. Seminar Nasional Aplikasi Teknologi Informasi 2010. Snati 2010. Aplikasi Mining Data Mahasiswa Dengan Metode Klasifikasi *Decision Tree* , 24-29.
- Sunjana, "Klasifikasi Data Nasabah Sebuah Asuransi Menggunakan Algoritma C4.5," Seminar Nasional Aplikasi Teknologi Informasi, pp. D31-D34, Juni 2010.
- Turban, E., dkk, 2005, *Decision Support Systems and Intelligent Systems*, Andi Offset
- T. H. Rahmadya and Prabowo P. W Herlawati, *Penerapan Data Mining Dengan Matlab*. Bandung : Rekayasa Sains: Andi, 2013.
- UCI, "Heart Disease Dataset," 2020. [Online]. Available: <https://archive.ics.uci.edu/ml/machine-learning-databases/statlog/heart/>. [Accessed: 01- Feb-2020].
- Undavia J N, Patel dan Dolia, "Comparison of Classification Algorithms to Predict Students' Post Graduation Course in Weka Environment," *International Journal of Advanced Research in Computer Science and Software Engineering*. vol. 3. no. 9. 2013.
- Virginia, Nancy (2020) Ketahui Gejala Penyakit Jantung Koroner Beserta Pengobatannya di <https://www.siloamhospitals.com/Contents/News-Events/Advertorial/2020/01/25/12/16/Ketahui-Gejala-Penyakit-Jantung-Koroner-Beserta-Pengobatannya> (diakses 25 April).
- WHO, "Cardiovascular diseases (CVDs)," 2015. [Online]. Available: https://www.who.int/cardiovascular_diseases/en/. [Accessed: 24-Des-2020].
- Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data Mining: Practical Machine Learning Tools and Techniques Third Edition*. Burlington: Morgan Kaufmann.