

**PREDIKSI RISIKO PENYAKIT JANTUNG KORONER
DENGAN METODE *ENSEMBLE* MENGGUNAKAN
ALGORITMA *NAIVE BAYES*, *DECISION TREE C4.5* DAN
REGRESI LOGISTIK BINER**

SKRIPSI

**Diajukan sebagai salah satu syarat memperoleh gelar Sarjana
di Jurusan Matematika Fakultas MIPA**

Oleh:

OLIVIA RIZKI INTAN MAULIA

NIM 08011181722056



**JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS SRIWIJAYA**

2021

LEMBAR PENGESAHAN

**PREDIKSI RISIKO PENYAKIT JANTUNG KORONER DENGAN
METODE *ENSEMBLE* MENGGUNAKAN ALGORITMA *NAIVE BAYES*,
DECISION TREE C4.5 DAN REGRESI LOGISTIK BINER**

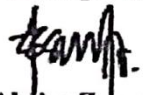
SKRIPSI

**Diajukan sebagai salah satu syarat memperoleh gelar
Sarjana di Jurusan Matematika Fakultas MIPA**

Oleh:

**OLIVIA RIZKI INTAN MAULIA
NIM. 08011181722056**

Pembimbing Pembantu



**Des Alwine Zavanti, M.Si
NIP.197012041998022001**

Indralaya, September 2021

Pembimbing Utama



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

Mengetahui,

Ketua Jurusan Matematika



HALAMAN PERSEMBAHAN

MOTTO

“Berusahalah! Jangan mudah menyerah dengan apapun yang terjadi. Tuhan selalu ada bersamamu, karena Tuhan Maha Tahu tentangmu”

**“Ia yang mengerjakan lebih dari apa yang dibayar pada suatu saat akan dibayar lebih dari apa yang ia kerjakan”
-Napoleon Hill**

Skripsi ini kupersembahkan kepada:

- 1. Allah SWT**
- 2. Bapak dan Ibu**
- 3. Saudaraku**
- 4. Keluarga Besar**
- 5. Dosen**
- 6. Almamater**
- 7. Sahabat dan Temanku**

KATA PENGANTAR

Assalamualaikum warahmatullahi wabarakatuh

Puji syukur kehadiran Tuhan Yang Maha Esa atas limpahan rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi dengan judul “**Prediksi Risiko Penyakit Jantung Koroner dengan Metode *Ensemble Majority Vote* Menggunakan Algoritma *Naive Bayes*, *Decision Tree C4.5* dan *Regresi Logistik Biner*”**. Shalawat beserta salam semoga selalu tercurahkan kepada baginda kita Nabi Muhammad SAW beserta keluarga, sahabat dan para pengikutnya hingga akhir zaman.

Penulis sadar dalam proses penyusunan skripsi ini masih banyak kekurangan, serta banyaknya rintangan dalam mengerjakannya. Namun dengan kesabaran dan ketekunan yang dilandasi rasa tanggung jawab sehingga penulis dapat menyelesaikan skripsi ini.

Penulis juga mengucapkan banyak terima kasih yang sebesar-besarnya kepada orang tua tercinta, yaitu **Bapak Juwanto** dan **Ibu Wiwin** yang tidak pernah lupa mendoakan, merawat dan memberikan kasih sayang, semangat serta restunya kepada penulis dan memberikan dukungan penuh atas apa yang penulis pilih. Penulisan skripsi ini tidak terlepas dari bantuan berbagai pihak, oleh karena itu penulis ingin mengucapkan terima kasih kepada:

1. **Kemenristekdikti** selaku pemberi beasiswa bidikmisi selama penulis menempuh perkuliahan sehingga penulis dapat merasakan menjadi mahasiswa dan menyelesaikan jenjang Strata Satu di Jurusan Matematika Universitas Sriwijaya.

2. **Bapak Drs. Sugandi Yahdin, M.M** selaku Ketua Jurusan Matematika FMIPA Universitas Sriwijaya atas ilmu yang telah diberikan.
3. **Ibu Dr. Dian Cahyawati Sukanda, M.Si** selaku Sekretaris Jurusan yang telah membantu dalam proses administrasi pendaftaran seminar, serta ilmu yang telah diberikan.
4. **Ibu Dr. Yulia Resti, M.Si** selaku Dosen Pembimbing Utama yang telah memberi masukan, saran, membimbing dan meluangkan waktu kepada penulis sehingga penulis dapat menyelesaikan skripsi dengan baik.
5. **Ibu Des Alwine Zayanti, M.Si** selaku Dosen Pembimbing Kedua yang telah memberikan masukan, saran dan membimbing penulis sehingga penulis dapat menyelesaikan skripsi dengan baik.
6. **Ibu Novi Rustiana Dewi, M.Si** selaku Dosen Pembahas I yang telah memberikan masukan, dan saran sehingga penulis dapat menyelesaikan skripsi dengan baik.
7. **Ibu Endang Sri Kresnawati, M.Si** selaku Dosen Pembahas II yang telah memberikan masukan, dan saran sehingga penulis dapat menyelesaikan skripsi dengan baik.
8. **Bapak Dr. Bambang Suprihatin M.Si** selaku Dosen Pembimbing Akademik serta Ketua Seminar skripsi yang telah memberikan arahan, saran, membimbing penulis serta mengatur jalannya seminar sehingga dapat berjalan dengan baik.
9. **Ibu Evi Yuliza, M.Si** selaku Sekretaris Seminar skripsi yang telah membantu dan memberikan catatan masukan yang sangat berguna bagi penulis.

10. **Seluruh Dosen di Jurusan Matematika FMIPA UNSRI** atas ilmu yang telah diberikan selama penulis menempuh pendidikan di Jurusan Matematika.
11. **Bapak Irwansyah dan Ibu Hamida** yang telah banyak membantu penulis dalam hal administrasi di Jurusan Matematika FMIPA UNSRI.
12. Saudara saya yang saya cintai **Tiffani, Dion, Menik, beserta para sepupu dan keponakkan** saya atas semangat yang diberikan kepada penulis.
13. Kepada (alm) **Bapak Somingan dan Ibu Yuyun Susilawati** yang telah banyak memberikan kasih sayang, dukungan secara material, dan selalu mendoakan penulis.
14. Tim Hebat, para sahabat seperjuangan **Abu, Agung, Yudha, Rendy, Shohif, Wawan, Mega, Azizah, Tesya, Kalimah, Muflhika** untuk saling menguatkan, mengajarkan, memberikan kesan tak terlupakan dan memberikan bantuannya selama proses penyelesaian skripsi ini.
15. **Sahabat Kareot Jecson, Apri, Kariah, Ririz, Eca** yang telah menemani dari awal maba dan selalu memberikan semangat pada penulis, serta teman-teman angkatan 17 yang tidak bisa diucapkan satu persatu.
16. **Sahabat BEM KM FMIPA Megi, Dian, Feni, Sisi, Eka, Juju, Echak, Jeni, Kak Hikma, Kak Mei, Kak Oki, Indra, Ramdan, Epan, Apresi, Rama, Seluruh Keluarga Eksternal** yang telah memberikan pengalaman berharga selama berorganisasi.
17. **Sahabat IKMS Yoga, Halim, Satria, Nurul, Ismi** yang telah memberikan pengalaman berharga selama berorganisasi.

18. **Kakak tingkat 2016** yang telah memberikan ilmu dan pengalamannya serta **adik tingkat angkatan 2018 dan 2019** yang telah memberikan dukungan dan doa serta semangat.
19. **Sahabat Dirot Dina, Rini dan Toiba** yang selalu memberikan doa dan support dari jauh.
20. **Sahabat SMA Juantri, Intan, Rara, Kintan, dan Thania** yang selalu memberikan dukungan dan semangat kepada penulis.
21. **Semua Pihak** yang telah membantu penulis dalam menyelesaikan skripsi ini yang tidak dapat disebutkan satu persatu.

Indralaya, 2021

Olivia Rizki Intan Maulia
NIM.08011181722056

**PREDICTION OF CORONARY HEART DISEASE RISK USING
ENSEMBLE METHOD USING NAIVE BAYES, DECISION TREE C4.5
AND BINARY LOGISTIC REGRESSION ALGORITHM**

By:

Olivia Rizki Intan Maulia

08011181722056

ABSTRACT

Coronary heart disease is one of the causes with high amount of death in Indonesia. Coronary heart disease is a coronary atherosclerotic disease that causes narrowing of blood vessels. Therefore, there are many who conduct research on coronary heart disease both on a large and small scale which is carried out by a classification process using a certain algorithm. This research aims to predict the risk of heart disease using the Ensemble method by combining the three classification algorithms. The data used in this study has 16 variables with a total data of 4238 data. The classification prediction uses the Ensemble Majority Vote method by combining the Naive Bayes algorithm, Decision Tree C4.5 and Binary Logistics Regression. The results of this study indicate that the prediction of the risk of heart disease using the Ensemble Majority Vote method obtains an accuracy rate of 84.79%, a precision of 86.01%, and a recall of 97.91%.

Keywords: Coronary Heart Disease, Ensemble Majority Vote, Naive Bayes, Decision Tree, Binary Logistics Regression.

**PREDIKSI RISIKO PENYAKIT JANTUNG KORONER DENGAN
METODE *ENSEMBLE* MENGGUNAKAN ALGORITMA *NAIVE BAYES*,
DECISION TREE C4.5 DAN REGRESI LOGISTIK BINER**

Oleh:

Olivia Rizki Intan Maulia

08011181722056

ABSTRAK

Penyakit jantung koroner merupakan salah satu penyakit yang menjadi penyebab kematian tertinggi di Indonesia. Penyakit jantung koroner adalah penyakit aterosklerotik koroner yang menyebabkan penyempitan pembuluh darah. Oleh karena itu banyak yang melakukan penelitian terhadap penyakit jantung koroner baik dalam skala jumlah besar maupun kecil yang dilakukan dengan proses klasifikasi menggunakan algoritma tertentu. Tujuan penelitian ini adalah memprediksi resiko penyakit jantung menggunakan metode *Ensemble* dengan menggabungkan ketiga algoritma pengklasifikasian. Data yang digunakan dalam penelitian memiliki 16 variabel dengan jumlah data sebesar 4238 data. Prediksi klasifikasi menggunakan metode *Ensemble Majority Vote* dengan menggabungkan algoritma *Naive Bayes*, *Decision Tree C4.5* dan Regresi Logistik Biner. Hasil penelitian ini menunjukkan bahwa prediksi risiko penyakit jantung dengan metode *Ensemble Majority Vote* memperoleh hasil tingkat *accuracy* sebesar 84,79%, *precision* sebesar 86,01%, dan *recall* sebesar 97,91%.

Kata Kunci: Penyakit Jantung Koroner, *Ensemble Majority Vote*, *Naive Bayes*, *Decision Tree*, Regresi Logistik Biner.

DAFTAR ISI

HALAMAN JUDUL	i
LEMBAR PENGESAHAN	ii
HALAMAN PERSEMBAHAN	iii
KATA PENGANTAR.....	iv
ABSTRACT	viii
ABSTRAK	ix
DAFTAR ISI.....	x
DAFTAR TABEL	xii
DAFTAR GAMBAR.....	xiii
DAFTAR LAMPIRAN	xiv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	5
1.3 Batasan Masalah.....	5
1.4 Tujuan Penelitian.....	6
1.5 Manfaat Penelitian.....	6
BAB II TINJAUAN PUSTAKA.....	8
2.1 Penyakit Jantung Koroner (PJK).....	8
2.2 Gejala PJK.....	8
2.3 Faktor Risiko PJK	9
2.4 Data Mining.....	9
2.5 <i>Machine Learning</i>	10
2.6 Klasifikasi.....	11
2.7 Metode <i>Ensemble</i>	12
2.8 Peluang.....	12
2.9 Algoritma <i>Naive Bayes</i>	13
2.10 <i>Laplace Smoothing</i>	15
2.11 Algoritma <i>Decision Tree</i>	16
2.11.1 Algoritma C4.5	16
2.12 Analisis Regresi.....	18
2.12.1 Analisis Regresi Logistik Biner	18
2.12.2 Pendugaan Parameter	19
2.12.3 Uji Serentak.....	20
2.12.4 Uji Hipotesis Parsial.....	21
2.12.5 Uji Kesesuaian Model	22
2.12.6 <i>Odds Ratio</i>	22
2.13 Confusion Matrix	23
BAB III METODOLOGI PENELITIAN	25
3.1 Tempat.....	25
3.2 Waktu	25
3.3 Metode Penelitian.....	25
BAB IV HASIL DAN PEMBAHASAN	28
4.1 Deskripsi Data	29
4.2 Diskritisasi Data	29

4.3 Ringkasan Data	31
4.4 Partisi Data	32
4.5 Algoritma <i>Naive Bayes</i>	33
4.6 Algoritma <i>Decision Tree</i>	40
4.7 Analisis Regresi Logistik	61
4.7.1 Uji Serentak.....	61
4.7.2 Uji Parsial.....	63
4.7.3 Uji Kesesuaian Model	66
4.7.4 <i>Odds Ratio</i>	67
4.7.5 Model Terbaik	69
4.7.6 Probabilitas.....	69
4.8 Metode Ensemble	71
4.9 Analisis Hasil	74
BAB V KESIMPULAN DAN SARAN	75
5.1 Kesimpulan.....	75
5.2 Saran.....	75
DAFTAR PUSTAKA	76
LAMPIRAN.....	79

DAFTAR TABEL

Tabel 2.1 <i>Confusion Matrix</i>	23
Tabel 4.1 Keterangan Variabel	28
Tabel 4.2 Diskritisasi Data	30
Tabel 4.3 Ringkasan data Variabel dependen	31
Tabel 4.4 Ringkasan data Variabel Independen	31
Tabel 4.5 Data <i>Train</i>	32
Tabel 4.6 Data <i>Test</i>	33
Tabel 4.7 Probabilitas Likelihood Variabel Independen	35
Tabel 4.8 Prediksi Algoritma <i>Naive Bayes</i>	39
Tabel 4.9 <i>Confusion Matrix</i> Algoritma <i>Naive Bayes</i>	39
Tabel 4.10 Prediksi Algoritma <i>Decision Tree</i>	59
Tabel 4.11 <i>Confusion Matrix</i> Algoritma <i>Decision Tree</i>	60
Tabel 4.12 Tabel Estimasi Model Regresi Logistik Biner	62
Tabel 4.13 Hasil Uji Serentak	62
Tabel 4.14 Hasil Uji Parsial	64
Tabel 4.15 Uji Hosmer dan Lemeshow	67
Tabel 4.16 Prediksi Algoritma Regresi Logistik Biner	70
Tabel 4.17 <i>Confusion Matrix</i> Algoritma Regresi Logistik Biner	71
Tabel 4.18 Metode <i>Ensemble</i>	72
Tabel 4.19 <i>Confusion Matrix</i> Metode <i>Esemble</i>	73
Tabel 4.20 Perbandingan Tingkat Ketepatan Prediksi	74

DAFTAR GAMBAR

Gambar 4.1 Pohon Keputusan.....	58
---------------------------------	----

DAFTAR LAMPIRAN

Lampiran 1 Hasil Pohon Keputusan <i>Decision Tree</i>	79
Lampiran 2 Tabel Chi Square	107
Lampiran 3 Regresi Logistik Biner dengan bantuan Software R	108
Lampiran 4 Hasil Prediksi.....	109

BAB I

PENDAHULUAN

1.1 Latar Belakang

Organ manusia yang berfungsi sebagai pemompa darah ke seluruh tubuh melalui pembuluh darah disebut jantung. Saat pembuluh darah mengalami masalah penyempitan, maka akan terjadi gangguan pada fungsi dari jantung tersebut sehingga mengakibatkan penyakit jantung (Hidayatul et al., 2018). Penyakit jantung sendiri memiliki banyak jenisnya, salah satunya adalah penyakit jantung koroner. Banyak yang beranggapan bahwa penyakit jantung koroner merupakan penyakit yang mematikan nomor dua setelah penyakit stroke (Tiani et al., 2015). Menurut WHO dari 17,5 juta kematian, 7,4 juta disebabkan oleh penyakit jantung koroner. Di Indonesia sendiri prevalensi penyakit jantung koroner berdasarkan diagnosis dokter sebesar 0,5% dan berdasarkan gejala dokter sebesar 1,5% (Sari et al., 2018).

Banyak faktor yang menyebabkan seseorang dapat menderita penyakit jantung koroner, dalam hal ini terbagi menjadi dua bagian faktor penyebab terjadinya penyakit jantung koroner yaitu faktor yang dapat diubah dan tidak dapat diubah (Patriyani dan Purwanto, 2016). Jenis kelamin, usia, riwayat penyakit dalam keluarga termasuk dalam faktor yang tidak dapat diubah. Sedangkan peningkatan lipid serum, tip kepribadian, merokok, tingkat glukosa, gaya hidup yang salah, stress psikologis, dan hipertensi obesitas sentral, kurang olahraga, diabetes, HDL-C rendah, LDL-C tinggi, faktor pembekuan termasuk dalam faktor yang dapat diubah (Poulter, 1999).

Pada penelitian ini, akan dilakukan proses klasifikasi untuk memprediksi risiko penyakit jantung koroner menggunakan proses data mining dengan machine learning metode klasifikasi. Metode yang digunakan dalam penelitian ini adalah Metode *Ensemble* yaitu dengan melakukan penggabungan algoritma diantaranya *Naive Bayes*, *Decision Tree* dan *Regresi Logistik Biner*.

Metode *Ensemble* merupakan gabungan dari beberapa algoritma sehingga kinerja yang diperoleh lebih baik dibandingkan hanya menggunakan satu algoritma saja (Dierrerich, 2001). Algoritma *Naive Bayes* adalah algoritma yang paling banyak digunakan dalam proses klasifikasi baik dalam bentuk angka ataupun kata (Maricar dan Pramana, 2019). Termasuk salah satu algoritma yang ditemukan oleh ilmuwan Inggris bernama Thomas Bayes, dikenal dengan Teorema Bayes dikarenakan metode ini memprediksi peluang di masa yang akan datang berdasarkan pengalaman yang ada sebelumnya (Sabransyah et al., 2017).

Decision Tree dapat dikatakan diagram alir yang berbentuk seperti pohon dimana setiap cabang dari pohon tersebut menyatakan hasil dari pengujian, suatu atribut dalam pohonnya disebut node. Node yang paling atas disebut sebagai *root node* atau node akar (Kasih, 2019).

Regresi Logistik Biner suatu algoritma yang bertujuan untuk menganalisa hubungan antara satu variabel dependen dengan variabel independen, dimana variabel dependennya bersifat dikotomi yaitu bernilai 0 dan 1 untuk menyatakan sebuah karakteristik atau label dimana 0 biasanya ketidakbenaran dan 1 biasanya sebaliknya. Regresi logistik biner juga dapat digunakan untuk memprediksi hasil klasifikasi dengan menggunakan model yang telah diperoleh (Rajagukguk et al., 2015).

Ada beberapa penelitian yang menerapkan metode Ensemble, *Naive Bayes Decision Tree*, dan Regresi Logistik Biner. Penelitian Indriati dan Kusyanti pada tahun 2018, melakukan penelitian dengan metode ensemble classifier untuk mendeteksi jenis ADHD pada anak usia dini. Algoritma yang digunakan adalah gabungan dari metode *K-nearest Neighbor*, *Fuzzy KNN*, *Neighbor Weighted K-Nearest Neighbor* (NWKNN) dan sebagainya. Hasil akurasi yang diperoleh dengan menggabungkan algoritma tersebut adalah 95% (Indriati dan Kusyanti, 2019).

Penelitian selanjutnya oleh Wibisono dan Fahrurrozi dalam melakukan perbandingan algoritma klasifikasi dalam pengklasifikasian data penyakit jantung koroner. Metode klasifikasi yang umum digunakan antara lain *Naive Bayes*, *K-Nearest Neighbor*, *Decision Tree* dan *Random Forest*. Data yang digunakan dalam penelitian tersebut adalah 300 dataset penyakit jantung koroner. Dari hasil penelitian diperoleh nilai ketepatan menggunakan algoritma *naive bayes* dan *decision tree* memiliki tingkat akurasi sebesar 80,33% (Wibisono dan Fahrurrozi, 2019).

Penelitian lain yang dilakukan oleh Fatmawati (2016), dalam penelitiannya melakukan perbandingan algoritma klasifikasi data mining untuk penyakit diabetes menggunakan Algoritma *Naive bayes* dan *Decision Tree*. Data yang digunakan dalam penelitian adalah hasil pemeriksaan terhadap 768 orang, 500 orang tidak terdeteksi penyakit diabetes dan 268 orang terdeteksi menderita penyakit diabetes. Ketepatan yang diperoleh dengan menggunakan Algoritma *Naive Bayes* adalah 75,13% dan menggunakan Algoritma *Decision Tree* adalah 73,30% (Fatmawati, 2016).

Penelitian lain yang dilakukan oleh Lishania dan tim (2019) melakukan penelitian dengan cara membandingkan Algoritma *Naive Bayes* dan *Decision Tree* untuk melakukan klasifikasi pada penderita penyakit *stroke* di RSUD Abdul Wahab Sjahranie Samarinda. Data yang digunakan dalam penelitian terdiri dari 156 data yang dibagi menjadi 90% data *train* atau sebanyak 140 sampel data dan data *test* sebesar 10% atau sebanyak 16 sampel. Hasil yang diperoleh pada penelitian ini dengan menggunakan Algoritma *Decision Tree* tingkat akurasi yang diperoleh hanya sebesar 87,5% dan menggunakan Algoritma *Naive Bayes* tingkat akurasi yang diperoleh sebesar 81,25% (Lishania et al., 2019).

Penelitian yang dilakukan oleh Attie dan tim (2008) melakukan Analisis Faktor Risiko Penyakit Jantung Koroner Menggunakan Metode Regresi Logistik dan Chaid, data yang digunakan dalam penelitian ini adalah data sekunder dari hasil penelitian yang dilakukan di Rumah Sakit Umum Dr. Wahidin Sudirohusodo Makassar tahun 2004-2007 terhadap pasien yang dirawat di rumah sakit tersebut yang datang berobat ke poliklinik penyakit dalam yang dinyatakan menderita penyakit jantung koroner (PJK) dan yang tidak menderita PJK dengan jumlah pasien sebanyak 827 orang. Hasil yang diperoleh dengan tingkat ketepatan *sensitivity* sebesar 74,7%, nilai *specificity* sebesar 59,2% dan akurasi adalah sebesar 67,2% (Atti et al., 2008).

Penelitian-penelitian lainnya yang melihat bahwa Algoritma *Naive Bayes*, *Decision tree C4.5* dan *Regresi Logistik Biner* mengklasifikasikan dengan ketepatan yang baik. Oleh karena itu penulis mencoba melakukan penelitian untuk mencari prediksi terbaik menggunakan Metode *Ensemble*. Jenis Metode *Ensemble* yang digunakan adalah dengan cara *vote*. *Vote* adalah mengambil prediksi dengan

suara atau nilai terbanyak dari berbagai model prediksi dalam memprediksi hasil dari proses klasifikasi. Algoritma yang digunakan dalam memprediksi data adalah *Naive Bayes*, *Decision tree* dan Regresi Logistik Biner. Hasil prediksi terakhir akan diperoleh menggunakan Metode *Ensemble* sehingga menghasilkan nilai *Accuracy*, *Precision* dan *Recall* yang tepat untuk klasifikasi data penyakit jantung koroner.

1.2 Rumusan Masalah

Adapun rumusan masalah yang di pada penelitian ini adalah :

1. Bagaimana hasil prediksi risiko penyakit jantung koroner menggunakan Metode *Ensemble* dengan penggabungan algoritma *Naive Bayes*, *Decision Tree C4.5* dan Regresi Logistik Biner?
2. Berapa tingkat ketepatan dalam memprediksi risiko penyakit jantung koroner menggunakan Metode *Ensembl* dengan algoritma *Naive Bayes*, *Decision Tree C4.5* dan Regresi Logistik Biner ?

1.3 Batasan Masalah

Adapun pembatasan masalah pada penelitian ini adalah:

1. Data *Cardiovascular study Dataset* dari *studi kardiovaskular* yang sedang berlangsung pada penduduk kota Framingham, Massachusetts tahun 2020. Data terdiri dari 16 variabel diantaranya adalah jenis kelamin, umur, *education*, *currentsmoker*, *cigs per days*, *bpmeds*, *prevalenstroke*, *prevalenthyp*, *diabetes*, *totchol*, *sysbp*, *diabp*, *bmi*, *glucosa*, *heart rate* dan *tenyearchd*.
2. Data yang digunakan sebanyak 4238 observasi, validasi data dengan rasio data train sebesar 80% dan data test 20%.

3. Tingkat ketepatan klasifikasi dibatasi oleh nilai *Accuracy*, *Precision* dan *Recall*.

1.4 Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah :

1. Memprediksi risiko penyakit jantung koroner menggunakan Metode *Ensemble* dengan algoritma *Naive Bayes*, *Decision Tree C4.5* dan Regresi Logistik Biner.
2. Menghitung tingkat ketepatan dalam memprediksi penyakit jantung koroner menggunakan metode *Ensemble* dengan algoritma *Naive Bayes*, *Decision Tree C4.5* dan Regresi Logistik Biner.

1.5 Manfaat Penelitian

Adapun manfaat dari penelitian ini adalah:

1. Dapat membantu dan mempermudah tenaga medis terkait informasi dalam memprediksi penyakit jantung.
2. Sebagai bahan referensi untuk penelitian yang juga membahas terkait prediksi risiko penyakit jantung.
3. Sebagai media pembelajaran penulis dan pembaca dalam memprediksi menggunakan metode *Naive bayes*, *Decision Tree C4.5* dan Regresi Logistik Biner.
4. Menambah pengetahuan bagi penulis dan pembaca tentang prediksi penyakit jantung.

DAFTAR PUSTAKA

- Ahmed, H., Abd-el ghany, S. F., Youn, E.M.G., Omran, N.F., & Ali. A.A., (2019). Stroke using Distributed Machine Learning Based on Apache Spark. *International Journal of Advanced Science and Techology*, 28(15), 89-97.
- Ahmed, K., & Jesmin, T. (2014). Comparative Analysis Of Data Mining Classification Algorithms in Type-2 Diabetes Prediction Data Using Weka Approach. *International Journal of Science and Engineering*, 7(2), 155–160.
- Atti, A., Sunarlim, B., & Syafitri, U. D. (2008). Analisis faktor risiko penyakit jantung koroner menggunakan metode regresi logistik dan chaid. *Jurnal Matematika Statistika*, 8(9), 138–148.
- Aulia, W. (2018). Sistem Pakar Diagnosa Penyakit Jantung Koroner Dengan Metode Probabilistic Fuzzy Decision Tree. *Jurnal Sains Dan Informatika*, 4(2), 106.
- Dietterich, T. G. (2001). Ensemble learning methods. *Handbook of Brain Theory and Neural Networks*, M.A. Arbib (ed.) (2nd ed). Cambridge: MIT Press.
- Fatmawati. (2016). Perbandingan Algoritma Klasifikasi Data Mining Untuk Model C4.5 Dan Naive Bayes Untuk Prediksi Penyakit Diabetes. *Jurnal Techno Nusa Mandiri*, 1(3), 137.
- Ghani, L., Susilawati, M. D., & Novriani, H. (2016). Faktor Risiko Dominan Penyakit Jantung Koroner di Indonesia. *Buletin Penelitian Kesehatan*, 44(3), 153–164.
- Hendrian, S. (2018). Algoritma Klasifikasi Data Mining Untuk Memprediksi Siswa Dalam Memperoleh Bantuan Dana Pendidikan. *Faktor Exacta*, 11(3), 266–274.
- Hidayatul, S., Aini, A., Sari, Y. A., & Arwan, A. (2018). *Seleksi Fitur Information Gain untuk Klasifikasi Penyakit Jantung Menggunakan Kombinasi Metode K-Nearest Neighbor dan Naïve Bayes*. 2(9), 2546–2554.
- Hosmer, D.W. dan S. Lemeshow, (2000) *Applied Logistic Regression*. Second Edition, John Willey & Sons, New York.
- Indriati, & Kusyanti, A. (2019). Metode Ensemble Classifier untuk Mendeteksi Jenis Attention Deficit Hyperactivity Disorder (SDHD) pada Anak Usia Dini. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 6(3), 301.
- Kareem, I. A., & Duaimi, M. G. (2014). Improved Accuracy for Decision Tree Algorithm Based on Unsupervised Discretization. *International Journal of Computer Science and Mobile Computing*, 36(6), 176–183.
- Kasih, P. (2019). Pemodelan Data Mining Decision Tree Dengan Classification Error Untuk Seleksi Calon Anggota Tim Paduan Suara. *Innovation in Research of Informatics (INNOVATICS)*, 1(2), 63–69.
- Lishania, I., Goejantoro, R., & Nasution, Y. N. (2019). Perbandingan Klasifikasi

- Metode Naive Bayes dan Metode Decision Tree Algoritma (J48) pada Pasien Penderita Penyakit Stroke di RSUD Abdul Wahab Sjahranie Samarinda. *Jurnal Eksponensial*, 10(2), 135–142.
- Listiowarni, I. (2019). Implementasi Naïve Bayesian dengan Laplacian Smoothing untuk Peminatan dan Lintas Minat Siswa SMAN 5 Pamekasan. *Jurnal Sisfokom (Sistem Informasi Dan Komputer)*, 8(2), 124.
- Maricar, M. A., & Dian Pramana. (2019). Perbandingan Akurasi Naïve Bayes dan K-Nearest Neighbor pada Klasifikasi untuk Meramalkan Status Pekerjaan Alumni ITB STIKOM Bali. *Jurnal Sistem Dan Informatika (JSI)*, 14(1), 16–22.
- Noviyanto. (2020). Penerapan Data Mining Dalam Mengelompokkan Jumlah Kematian. *Paradigma-Jurnal Informatika Dan Komputer*, 22(2).
- Patil, T.R., & Sherekar, S. . (2019). Performance Analysis of Naive Bayes and J48 Classification Algorithm for Data Classification. *International Journal of Computer Science and Applications*, 7(2), 88–91.
- Patriyani, R. E. H., & Purwanto, D. F. (2016). Faktor Dominan Risiko Terjadinya Penyakit Jantung Koroner (PJK). (*Jkg*) *Jurnal Keperawatan Global*, 1(1),
- Poulter, N. (1999). Coronary heart disease is a multifactorial disease, *American Journal of Hypertensions*, 7(10 SUPPL. 1), 92-95.
- Priyam, A., Gupta, R., Rathee, A., & Srivastava, S. (2013). Comparative Analysis of Decision Tree Classification Algorithms. *International Journal of Current Engineering and Technology*, 3, 334–337.
- Rajagukguk, N., Ispriyanti, D., & Wilandari, Y. (2015). Perbandingan Metode Klasifikasi Regresi Logistik Biner dan Naive Bayes Pada Status Pengguna KB Di Kota Tegal Tahun 2014. *Jurnal Gaussian*, 4, 365–374.
- Ratnawati, E., & Sunarko. (2008). Evaluasi Kinerja Fasilitas Iradiasi Sistem Rabbit Menggunakan Bahan Acuan Standard Dengan Metode AAN. *Buletin Pengelolaan Reaktor Nuklir*, 5(2), 49–55.
- Rosandy, T. (2016). Perbandingan Metode Naive Bayes Classifier dengan Metode Decision Tree Untuk Menganalisa Kelancaran Pembiayaan. *Jurnal TIM Darmajaya*, 02(01), 52–62.
- Sabransyah, M., Nasution, Y. N., & Amijaya, F. D. T. (2017). Aplikasi Metode Naive Bayes dalam Prediksi Risiko Penyakit Jantung. *Jurnal EKSPONENSIAL*, 8(2), 111–118.
- Sari, D. M., Azrimaidaliza, A., & Purnakarya, I. (2010). Faktor resiko kolesterol total pasien penyakit jantung koroner di rumah sakit achmad mochtar bukittinggi. *Jurnal Kesehatan Masyarakat*, 4(2), 77–81.
- Sari, E. A., Sari, S. P., & Pratiwi, S. H. (2018). Community Self-Efficacy of Coronary Heart Disease Based on Characteristic Risk Factors. *NurseLine Journal*, 3(2), 65.
- Sindar, A., & Sitorus, M. (2020). Machine Learning Prediksi Karakter Pengguna

- Hastag (#) Bahasa Generasi Milenial Di Sosial Media. *Indonesian Journal of Applied Informaties*, 3, 68–70.
- Tampil, Y., Komaliq, H., & Langi, Y. (2017). Analisis Regresi Logistik Untuk Menentukan Faktor-Faktor Yang Mempengaruhi Indeks Prestasi Kumulatif (IPK) Mahasiswa FMIPA Universitas Sam Ratulangi Manado. *D’CARTESIAN*, 6(2), 56.
- Tiani, S., Wahjudi, P., & Wati, D. M. (2015). *Determinan Penyakit Jantung dan Pembuluh Darah Pasien Rawat Jalan Poli Jantung dan Poli Penyakit Dalam RSD dr . Soebandi Jember Cardiovascular and Internal Medicine Outpatients Clinic*. 3(3).
- Wahono, R. S., Supriyanto, C., Komputer, F. I., & Nuswantoro, U. D. (2015). Penanganan Fitur Kontinyu dengan Feature Discretization Berbasis Expectation Maximization Clustering untuk Klasifikasi Spam Email Menggunakan Algoritma ID3. *Journal of Intelligent Systems*, 1(2), 148–155.
- Walpole, R. E. (1993). *Pengantar Statistika* (3rd ed). Jakarta: Gramedia Pustaka Utama
- Wibisono, A. B., & Fahrurozi, A. (2019). Perbandingan Algoritma Klasifikasi Dalam Pengklasifikasian Data Penyakit Jantung Koroner. *Jurnal Ilmiah Teknologi Dan Rekayasa*, 24(3), 161–170.