

**KLASIFIKASI KEJADIAN HIPERTENSI
DENGAN METODE *SUPPORT VECTOR MACHINE* (SVM)
MENGUNAKAN DATA PUSKESMAS DI KOTA BANDAR LAMPUNG**

(Skripsi)

Oleh

**INDAH MARTIKA PASARIBU
NPM 1617051105**



**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2021**

**KLASIFIKASI KEJADIAN HIPERTENSI
DENGAN METODE *SUPPORT VECTOR MACHINE* (SVM)
MENGUNAKAN DATA PUSKESMAS DI KOTA BANDAR LAMPUNG**

Oleh

INDAH MARTIKA PASARIBU

Skripsi

**Sebagai Salah Satu Syarat untuk Mencapai Gelar
SARJANA ILMU KOMPUTER**

Pada

**Jurusan Ilmu Komputer
Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Lampung**



**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2021**

ABSTRAK

**KLASIFIKASI KEJADIAN HIPERTENSI
DENGAN METODE *SUPPORT VECTOR MACHINE* (SVM)
MENGUNAKAN DATA PUSKESMAS DI KOTA BANDAR LAMPUNG**

Oleh

INDAH MARTIKA PASARIBU

Hipertensi adalah keadaan seseorang mengalami peningkatan tekanan darah di atas nilai normal yang menyebabkan rasa sakit hingga kematian. Tekanan darah normal manusia adalah 120/80 mmHg. Penderita hipertensi tidak dapat disembuhkan, namun dapat dilakukan pencegahan dan pengontrolan. Permasalahan hipertensi selalu mengalami peningkatan di Indonesia. Dinas Kesehatan Kota Bandar Lampung mengemukakan bahwa penyakit hipertensi adalah penyakit yang selalu memasuki peringkat sepuluh penyakit terbanyak di Kota Bandar Lampung. Diagnosis penyakit hipertensi dilakukan secara manual dengan membutuhkan tenaga, materi, dan waktu yang banyak saat ini. Berdasarkan hal tersebut, terdapat sebuah gagasan untuk menerapkan bidang ilmu analisis *biomedical* data pada klasifikasi penyakit hipertensi dengan menggunakan metode *support vector machine* (SVM) di Kota Bandar Lampung.

Penelitian ini mengklasifikasi dan mengukur akurasi metode *support vector machine* pada hipertensi. Data berasal dari lima puskesmas di Kota Bandar Lampung tahun 2017 sampai 2019 dengan pembagian data *10-fold cross validation*. *Kernel* yang digunakan adalah *kernel* linear, *gaussian*, dan *polynomial*. Hasil dari penelitian ini adalah berhasil mengklasifikasi penderita hipertensi di Kota Bandar Lampung. Hasil analisis korelasi fitur tertinggi sebesar 0.90. Hasil klasifikasi menggunakan metode *support vector machine* mendapatkan akurasi tertinggi sebesar 99.78% pada *kernel gaussian*.

Kata kunci: Hipertensi, *Kernel*, *K-Fold Cross Validation*, Klasifikasi, *Support Vector Machine*.

ABSTRACT

CLASSIFICATION OF HYPERTENSION EVENTS WITH SUPPORT VECTOR MACHINE (SVM) METHOD USING HEALTH CENTERS DATA IN BANDAR LAMPUNG

By

INDAH MARTIKA PASARIBU

Hypertension is a condition in which a person experiences an increase in blood pressure above the normal value which causes pain and even death. Normal human blood pressure is 120/80 mmHg. Patients with hypertension cannot be cured, but prevention and control can be done. The hypertension cases is always increasing in Indonesia. The Bandar Lampung City Health Service stated that hypertension is a disease that always ranks in the top ten diseases in Bandar Lampung City. Diagnosis of hypertension is currently manually performed by requiring a lot of energy, materials, and time. Based on condition, there is an idea to apply the field of biomedical data analysis to help the diagnosing hypertension using the support vector machine (SVM) method in Bandar Lampung City. This study classifies and measures the accuracy of the support vector machine method in hypertension. The data comes from five health centers in Bandar Lampung City from 2017 to 2019 with 10-fold cross validation datasharing. The kernels used are linear, gaussian, and polynomial kernels. This study successfully classifies hypertension sufferers in Bandar Lampung City. The result of the highest feature correlation analysis is 0.90. The results of the classification using the support vector machine method get the highest accuracy, which is 99.78% on the gaussian kernel.

Keywords: Classification, Hypertension, Kernel, K-Fold Cross Validation, Support Vector Machine.

Judul Skripsi : **KLASIFIKASI KEJADIAN HIPERTENSI
DENGAN METODE *SUPPORT VECTOR
MACHINE (SVM)* MENGGUNAKAN DATA
PUSKESMAS DI KOTA BANDAR
LAMPUNG**

Nama Mahasiswa : **Indah Martika Pasaribu**

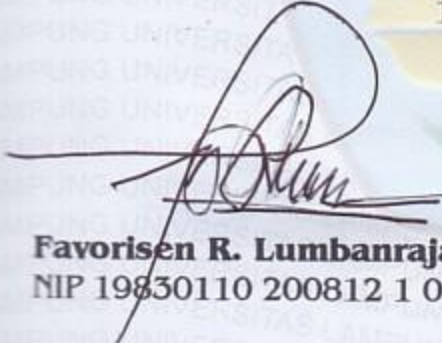
Nomor Pokok Mahasiswa : 1617051105

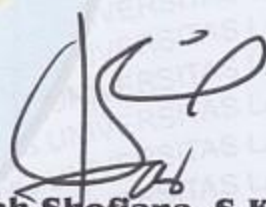
Program Studi : Ilmu Komputer

Fakultas : Matematika dan Ilmu Pengetahuan Alam

MENYETUJUI

1. Komisi Pembimbing


Favorisen R. Lumbanraja, Ph.D.
NIP 19830110 200812 1 002


Dewi Asiah Shofiana, S.Komp., M.Kom.
NIP 19950929 202012 2 030

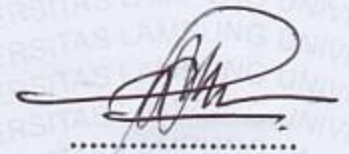
2. Ketua Jurusan Ilmu Komputer


Didik Kurniawan, S.Si., M.T.
NIP 19800419 200501 1 004

MENGESAHKAN

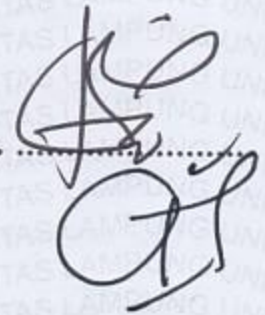
1. Tim Penguji

Ketua : **Favorisen R. Lumbanraja, Ph.D.**



.....

Sekretaris : **Dewi Asiah Shofiana, S.Komp., M.Kom.**



.....

Anggota : **Aristoteles, S.Si., M.Si.**

.....

2. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam



Dr. Eng. Surtipto Dwi Yuwono, S.Si., M.T.

NIP 19740705 200003 1 001

Tanggal Lulus Ujian Skripsi : 13 Agustus 2021

PERNYATAAN

Saya yang bertanda tangan di bawah ini, menyatakan bahwa skripsi saya yang berjudul “Klasifikasi Kejadian Hipertensi dengan Metode *Support Vector Machine* (SVM) Menggunakan Data Puskesmas di Kota Bandar Lampung” merupakan karya saya sendiri dan bukan karya orang lain. Semua tulisan yang tertuang di skripsi ini telah mengikuti kaidah penulisan karya ilmiah Universitas Lampung. Apabila dikemudian hari terbukti skripsi saya merupakan hasil penjiplakan atau dibuat orang lain, maka saya bersedia menerima sanksi berupa pencabutan gelar yang saya terima.

Bandar Lampung, 16 Agustus 2021




Indah Martika Pasaribu
NPM. 1617051105

RIWAYAT HIDUP



Penulis dilahirkan di Kisaran pada 22 Maret 1998, sebagai anak pertama dari empat bersaudara dari ayah yang bernama Parulian Pasaribu, S.Pd., M.Si. dan ibu bernama Taruli Martulus Sitorus S.Tr.Keb. Penulis menyelesaikan pendidikan formal pertama kali di TK Jaya Wijaya pada tahun 2004.

Pendidikan dasar di SD Negeri 010131 Pulau Rakyat Tua diselesaikan pada tahun 2010. Pendidikan menengah pertama di SMP Negeri 1 Pulau Rakyat diselesaikan pada tahun 2013, kemudian melanjutkan pendidikan di SMA Negeri 2 Pematang Siantar yang diselesaikan pada tahun 2016.

Pada tahun 2016 penulis melanjutkan pendidikan dan terdaftar sebagai mahasiswa Jurusan Ilmu Komputer Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung. Selama menjadi mahasiswa penulis melakukan beberapa kegiatan antara lain.

1. Menjadi anggota ABACUS (Anggota Baru Computer Science) Jurusan Ilmu Komputer Universitas Lampung pada periode 2016.
2. Menjadi anggota divisi Hubungan Masyarakat Unit Kegiatan Mahasiswa Kristen Universitas Lampung pada periode kepengurusan tahun 2017.

3. Menjadi ketua divisi Hubungan Masyarakat Unit Kegiatan Mahasiswa Kristen Universitas Lampung pada periode kepengurusan tahun 2018.
4. Menjadi asisten dosen di Jurusan Ilmu Komputer.
5. Melaksanakan kerja praktik di Puskesmas Aek Songsongan, Kecamatan Bandar Pulau, Kabupaten Asahan, Sumatera Utara pada Januari tahun 2019.
6. Melaksanakan kegiatan Kuliah Kerja Nyata (KKN) di Desa Suka Menanti, Kecamatan Bukit Kemuning, Kabupaten Lampung Utara, Lampung pada bulan Juni tahun 2019.

PERSEMBAHAN

Puji syukur kepada Tuhan Yesus Kristus atas berkat dan kasih-Nya telah memberikan kekuatan untuk dapat menyelesaikan skripsi ini.

Skripsi ini saya persembahkan kepada:

Kedua orang tua, Bapak dan Mama yang telah memberikan kasih sayang, doa dan dukungan moral maupun material, dan yang selalu mendidik anak-anaknya menjadi pribadi yang lebih baik. Terima kasih juga kepada ketiga adik yang memberikan semangat.

Keluarga Ilmu Komputer.

Almamater tercinta, Universitas Lampung.

MOTTO

“Takut akan Tuhan adalah permulaan pengetahuan.” (Amsal 1:7 a)

“Apapun juga yang kamu perbuat, perbuatlah dengan segenap hatimu seperti untuk Tuhan dan bukan untuk manusia.” (Kolose 3:23)

“Diberkatilah orang yang mengandalkan Tuhan, yang menaruh harapannya pada Tuhan!” (Yeremia 17:7)

“Janganlah hendaknya kamu kuatir tentang apa pun juga, tetapi nyatakanlah dalam segala hal keinginanmu kepada Allah dalam doa dan permohonan dengan ucapan syukur.” (Filipi 4:6)

“Janganlah takut, sebab Aku menyertai engkau, janganlah bimbang, sebab Aku ini Allahmu; Aku akan meneguhkan, bahkan akan menolong engkau; Aku akan memegang engkau dengan tangan kanan-Ku yang membawa kemenangan.”

(Yesaya 41:10)

“Marilah kepada Ku, semua yang letih lesu dan berbeban berat, Aku akan memberi kelegaan kepadamu.” (Matius 11:28)

“Dan apa saja yang kamu minta dalam doa dengan penuh kepercayaan, kamu akan menerimanya.” (Matius 21:22)

“This too shall pass.” (Anonim ; 2 Corinthians 4:17-18)

“Percayalah rancangan Tuhan Yesus selalu tepat dan indah. Terima kasih Bapa.

Amin.”

(Indah Martika Pasaribu)

UCAPAN TERIMA KASIH

Puji syukur kepada Tuhan Yesus Kristus, karena atas berkat dan penyertaan-Nya penulis dapat menyelesaikan skripsi yang berjudul “Klasifikasi Kejadian Hipertensi dengan Metode *Support Vector Machine* (SVM) Menggunakan Data Puskesmas di Kota Bandar Lampung”. Skripsi ini merupakan salah satu syarat untuk memperoleh gelar sarjana Ilmu Komputer di Universitas Lampung.

Dalam penyusunan skripsi ini penulis menyadari banyak mendapat bimbingan, dukungan, dan motivasi dari berbagai pihak yang dengan tulus telah membantu penulis untuk menyelesaikan skripsi ini. Penulis sangat bersyukur dan mengucapkan terima kasih sedalam-dalamnya kepada:

1. Kedua orang tua, Bapak dan Mama yang telah senantiasa memberikan doa, dukungan, pembelajaran, dan kasih sayang yang sangat luar biasa.
2. Intan Maystarni Pasaribu, Indrani Masprisa Pasaribu, dan Inchristo Manfrankly Jeremy Pasaribu yang selalu memberikan doa, dukungan, dan kasih sayang.
3. Bapak Favorisen R. Lumbanraja, Ph.D., selaku dosen pembimbing utama atas kesediaannya dan kesabarannya untuk memberikan dukungan, bimbingan, kritik, dan saran dalam proses penyelesaian skripsi.
4. Ibu Dewi Asiah Shofiana, S.Komp., M.Kom., selaku dosen pembimbing kedua atas kesediaannya dalam memberikan bimbingan, nasihat, kritik, dan saran selama proses pengerjaan skripsi.
5. Bapak Aristoteles, S.Si., M.Si., selaku dosen pembahas skripsi yang telah memberikan saran dan masukan dalam penyempurnaan penulisan skripsi.

6. Bapak Dr. Eng. Supto Dwi Yuwono, M.T. selaku Dekan FMIPA Universitas Lampung.
7. Bapak Didik Kurniawan, S.Si., M.T., selaku Ketua Jurusan Ilmu Komputer FMIPA Universitas Lampung.
8. Ibu Astria Hijriani, S.Kom., M.Kom., selaku Sekretaris Jurusan Ilmu Komputer FMIPA Universitas Lampung.
9. Bapak Tristiyanto, S.Kom., M.I.S., Ph.D., selaku dosen pembimbing akademik.
10. Bapak dan Ibu Dosen Jurusan Ilmu Komputer Universitas Lampung yang telah memberikan ilmu dan pengetahuan hidup selama penulis menjadi mahasiswa.
11. Ibu Ade Nora Maela, Pak Zainudin, dan Mas Nofal yang telah memudahkan segala urusan administrasi penulis di Jurusan Ilmu Komputer.
12. Teman-teman Kriskat (Kristen Katolik) 2016 yang telah senantiasa menjadi teman yang menemani keseharian penulis selama masa perkuliahan dengan canda dan tawa.
13. Teman-teman di Jurusan Ilmu Komputer Angkatan 2016 yang menjadi teman satu angkatan selama menjalankan masa studi di Universitas Lampung.
14. Teman-teman di fakultas MIPA yang menemani penulis.
15. Teman-teman di Unit Kegiatan Mahasiswa Kristen Universitas Lampung yang telah mengajarkan dalam berpelayanan dan berorganisasi selama kepengurusan dan kepanitiaan.
16. Teman-teman seperbimbingan yang telah menemani dan membantu untuk sama-sama belajar dalam menyusun sampai menyelesaikan skripsi ini.
17. Seluruh keluarga besar yang memberikan semangat dan dukungan kepada penulis.
18. Semua pihak yang secara langsung atau tidak langsung yang telah membantu menyelesaikan skripsi ini.

Penulis menyadari bahwa penulisan skripsi ini masih jauh dari kata sempurna karena masih terbatasnya pengetahuan, pengalaman, dan kemampuan penulis. Oleh karena itu penulis sangat menghargai dan mengharapkan saran dan kritik untuk penelitian ini sebagai bahan pertimbangan untuk karya tulis yang akan datang. Semoga skripsi ini dapat bermanfaat untuk semua pihak.

Bandar Lampung, 16 Agustus 2021

Indah Martika Pasaribu
NPM. 1617051105

DAFTAR ISI

	Halaman
DAFTAR ISI	i
DAFTAR GAMBAR	iii
DAFTAR TABEL	iv
DAFTAR KODE PROGRAM	v
I. PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah.....	4
1.4 Tujuan Penelitian.....	4
1.5 Manfaat Penelitian.....	4
II. TINJAUAN PUSTAKA	5
2.1 Penelitian Terdahulu.....	5
2.2 Uraian Landasan Teori	7
2.2.1 Hipertensi	7
2.2.2 <i>Machine Learning</i>	9
2.2.3 <i>Support Vector Machine (SVM)</i>	10
2.2.4 <i>Feature Transformation</i>	13
2.2.5 Analisis Korelasi Fitur... ..	13
2.2.6 <i>Cross Validation</i>	13

2.2.7	<i>Confusion Matrix</i>	16
III.	METODOLOGI PENELITIAN	19
3.1	Tempat Dan Waktu Penelitian	19
3.1.1	Tempat	19
3.1.2	Waktu	20
3.2	Alat Dan Data	21
3.2.1	Alat.....	21
3.2.2	Data.....	22
3.3	Metode.....	23
IV.	HASIL DAN PEMBAHASAN	27
4.1	<i>Feature Transformation</i>	27
4.2	Analisis Korelasi Fitur.....	31
4.3	Percobaan Model Klasifikasi.....	35
4.4	Pembahasan	43
V.	SIMPULAN DAN SARAN	45
5.1	Simpulan.....	45
5.2	Saran	46
	DAFTAR PUSTAKA	47

DAFTAR GAMBAR

Gambar	Halaman
1. <i>Prevalensi</i> hipertensi di Indonesia tahun 2007-2018 (Kementerian Kesehatan Republik Indonesia, 2018).....	2
2. Pendekatan konseptual pada metode <i>support vector machine</i> (Cortes & Vapnik, 1995).....	11
3. Proses <i>k-fold cross validation</i> (Lukito & Chrismanto, 2015)	15
4. Proses <i>leave-one-out cross validation</i> (Cawley & Talbot, 2003)	16
5. <i>Flowchart</i> tahap penelitian klasifikasi hipertensi.....	24
6. Grafik hasil prediksi menggunakan kernel <i>linear</i> , <i>gaussian</i> , dan <i>polynomial</i>	44

DAFTAR TABEL

Tabel		Halaman
1.	Klasifikasi Tekanan Darah (Triyanto, 2014).....	8
2.	Interpretasi Koefisien Korelasi (Iin & Sugiarto, 2012).....	14
3.	Klasifikasi Evaluasi Matriks (Boughorbel <i>et al.</i> , 2017).....	17
4.	<i>Gantt Chart</i> Waktu Pelaksanaan Penelitian Klasifikasi Hipertensi	21
5.	<i>Data Set</i> Penelitian Klasifikasi Hipertensi	23
6.	Contoh Data Hasil <i>Feature Transformation</i>	29
7.	Hasil Analisis Korelasi Fitur	32
8.	Pengelompokan dan Pemanggilan Data dengan Dua Kelas.....	36
9.	Hasil <i>Confusion Matrix</i> Menggunakan <i>Kernel Linear</i>	38
10.	Hasil Prediksi Model SVM Klasifikasi Menggunakan <i>Kernel Linear</i>	39
11.	Hasil <i>Confusion Matrix</i> Menggunakan <i>Kernel Gaussian</i>	40
12.	Hasil Prediksi Model SVM Klasifikasi Menggunakan <i>Kernel Gaussian</i>	41
13.	Hasil <i>Confusion Matrix</i> Menggunakan <i>Kernel Polynomial</i>	42
14.	Hasil Prediksi Model SVM Klasifikasi Menggunakan <i>Kernel Polynomial</i>	42

DAFTAR KODE PROGRAM

Kode Program		Halaman
1.	Menampilkan Hasil Analisis Korelasi Fitur.....	31
2.	<i>Import</i> Data dan Penganalisan Data.	36
3.	Pengelompokan dan Pemanggilan Data Berdasarkan Kelas ...	36
4.	Pembagian <i>Data Set</i>	37
5.	Membuat Model Klasifikasi.....	37
6.	Melakukan Prediksi dengan Model Klasifikasi.....	38
7.	Mencari Kinerja dari Hasil Prediksi	38

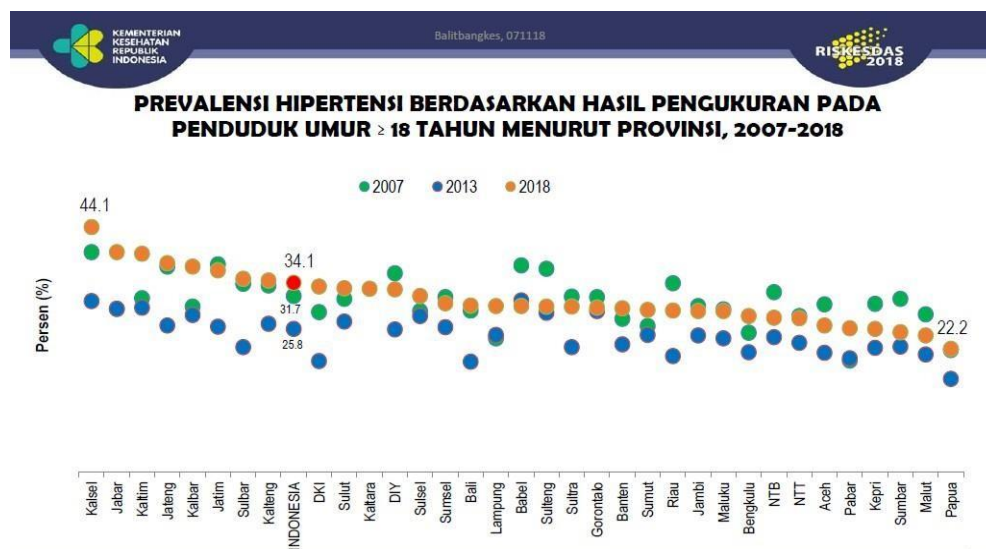
I. PENDAHULUAN

1.1 Latar Belakang

Hipertensi adalah keadaan seseorang mengalami peningkatan tekanan darah di atas nilai normal yang menyebabkan rasa sakit hingga kematian. Tekanan darah normal manusia adalah 120/80 mmHg (Herwati & Sartika, 2014). Hipertensi menyebabkan penderita mengalami risiko kerusakan mata, penyakit jantung (*cardiac*), kerusakan otak (*stroke*), pengerasan pembuluh darah (*atherosclerosis*), penggumpalan pembuluh darah (*arteriosclerosis*), dan penyakit ginjal (*renal*) (Triyanto, 2014). Hipertensi disebut *the silent killer* atau pembunuh gelap, karena beberapa penderita tidak menunjukkan gejala hipertensi (Susilo & Wulandari, 2011).

Hipertensi disebabkan oleh dua faktor, yaitu faktor keturunan dan pola hidup. Saat ini, penderita hipertensi yang disebabkan oleh faktor keturunan (genetika) belum diketahui letak kerusakan genetiknya. Penderita hipertensi oleh faktor pola hidup disebabkan karena banyak mengonsumsi garam, alkohol, merokok, obesitas, stres, dan diabetes (Rahajeng & Tuminah, 2009). Penderita hipertensi tidak dapat disembuhkan, namun dapat dilakukan pencegahan dan pengontrolan. Oleh karena itu, deteksi dini sangat diperlukan untuk meningkatkan perawatan kesehatan umum (Arda *et al.*, 2018).

Permasalahan hipertensi selalu mengalami peningkatan di Indonesia. Hasil Riset Kesehatan Dasar (Riskesdas) tahun 2018 menyatakan bahwa *prevalensi* penderita hipertensi adalah 34.1%. Pada grafik Gambar 1 digambarkan bahwa jumlah tersebut mengalami peningkatan sebesar 8,3% dari tahun 2013. *Prevalensi* penyakit hipertensi tahun 2013 adalah 25.8%. Hasil Riset Kesehatan Dasar (Riskesdas) tahun 2018 pada Gambar 1 menyatakan bahwa penderita hipertensi terbanyak berada di Provinsi Kalimantan Selatan (44,1%) dan penderita hipertensi paling sedikit berada di Provinsi Papua (22,2%). *Prevalensi* penyakit hipertensi dari tahun 2007 sampai tahun 2018 menyatakan bahwa penderita hipertensi di Provinsi Lampung selalu mengalami peningkatan. Data dari Dinas Kesehatan Kota Bandar Lampung dari tahun 2015 sampai tahun 2018 mengemukakan bahwa penyakit hipertensi adalah penyakit yang selalu memasuki peringkat sepuluh penyakit terbanyak di Kota Bandar Lampung.



Gambar 1. *Prevalensi* hipertensi di Indonesia tahun 2007-2018 (Kementerian Kesehatan Republik Indonesia, 2018).

Penelitian penyakit hipertensi yang berjudul “A Study on Comparative Performance of SVM Classifier Models with Kernel Functions in Prediction of Hypertension” yang disusun oleh (Samant & Rao, 2013) menggunakan metode *support vector machine* (SVM) untuk memprediksi probabilitas

terjadinya hipertensi esensial (HT) pada populasi pasien campuran. *Data set* yang digunakan terdiri dari pasien sehat dan diabetes dari rumah sakit Indian Institute of Technology Bombay (IITB), Mumbai, India. *Data set* tersusun dari 13 atribut yang berupa gejala. *Kernel* yang digunakan adalah linear, *quadratic*, *polyorder (order three)*, *multi layer perceptron (MLP)*, dan *radial basis function (RBF)*. Fungsi *kernel support vector machine* yang diuji menunjukkan akurasi yang cukup baik dalam prediksi penyakit hipertensi.

Diagnosis penyakit hipertensi dilakukan secara manual selama ini. Pasien akan diperiksa oleh tenaga medis secara rutin. Pasien diperiksa dengan menggunakan stetoskop dan tensimeter (Susilo & Wulandari, 2011). Hal ini akan membutuhkan tenaga, materi, dan waktu yang banyak. Berdasarkan hal tersebut, terdapat sebuah gagasan untuk menerapkan bidang ilmu analisis *biomedical data* pada klasifikasi penyakit hipertensi dengan menggunakan metode *support vector machine (SVM)* di Kota Bandar Lampung. Berdasarkan hasil penelitian (Osisanwo *et al.*, 2017) dikemukakan bahwa metode *support vector machine* adalah algoritme presisi dan paling akurat pada klasifikasi. *Support vector machine* adalah metode untuk menemukan fungsi pemisah (*classifier*) yang optimal dari dua kelas yang berbeda. Kelebihan SVM adalah proses komputasi yang cepat karena menggunakan *support vector* (Cortes & Vapnik, 1995).

1.2 Rumusan Masalah

Rumusan masalah dalam penelitian ini adalah bagaimana mengetahui nilai keakuratan klasifikasi kejadian hipertensi menggunakan metode *support vector machine (SVM)* di Kota Bandar Lampung.

1.3 Batasan Masalah

Batasan masalah penelitian ini adalah sebagai berikut:

1. Penelitian hanya menggunakan metode *support vector machine*.
2. Penggunaan metode *support vector machine* dengan *kernel* linear, *polynomial*, dan *gaussian*.
3. Pengukuran parameter menggunakan *confusion matrix* yang terdiri dari parameter *accuracy*, *sensitivity/recall*, dan *precision*.
4. Data berasal dari lima puskesmas di Kota Bandar Lampung tahun 2017 sampai 2019

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut:

1. Mengklasifikasi hipertensi dengan metode *support vector machine*.
2. Mengukur akurasi metode *support vector machine* pada hipertensi.

1.5 Manfaat Penelitian

Manfaat penelitian ini adalah sebagai berikut:

1. Implementasi model klasifikasi pada hipertensi.
2. Menambah kepaduan informasi untuk mengantisipasi hipertensi.

II. TINJAUAN PUSTAKA

2.1 Penelitian Terdahulu

Penelitian ini memerlukan penelitian sebelumnya untuk membandingkan hasil penelitian. Berikut adalah sumber penelitian terdahulu sebagai pendukung penelitian ini, yaitu:

2.1.1 Penelitian berjudul “*A Study on Comparative Performance of SVM Classifier Models with Kernel Functions in Prediction of Hypertension*” yang disusun oleh (Samant & Rao, 2013) menjadi sumber acuan pada penelitian ini. *Kernel* yang digunakan adalah linear, *quadratic*, *polyorder (order three)*, *multi layer perceptron (MLP)*, dan *radial basis function (RBF)*. *Data set* yang digunakan berasal dari rumah sakit Indian Institute of Technology Bombay (IITB), Mumbai, India. Data yang digunakan sekitar 20 tahun selama periode Januari 1995 sampai April 2005. Penggunaan *data set* tersebut sebanyak 981 data. Atribut yang digunakan dalam penelitian (Samant & Rao, 2013) yaitu usia, *systolic blood pressure (BP1)*, *diastolic blood pressure (BP2)*, *contraves 30 viscometer*, *serum proteins (SP)*, *serum albumin (SALB)*, *hemorheological (HR)*, *hematocrit (HCT)*, *serum cholesterol (SC)*, *red cell aggregation (RCA)*, *serum triglycerides (STG)*, *whole blood viscosity (CBV)*, dan *plasma viscosity (CPV)*. Penelitian tersebut melakukan empat klasifikasi pada *data set* yang sama. Akurasi klasifikasi pertama, kedua, dan ketiga

lebih tinggi dibandingkan dengan klasifikasi keempat. Hal ini disebabkan karena klasifikasi keempat menggunakan KNN-*imputed* untuk nilai yang hilang. Penelitian tersebut menunjukkan bahwa *kernel linier* pada klasifikasi ketiga adalah *kernel* yang terbaik dengan akurasi sebesar 88.4%.

2.1.2 Penelitian “Klasifikasi Risiko Hipertensi Menggunakan Metode *Neighbor Weighted K-Nearest Neighbor* (NWKNN)” yang disusun oleh (Yudha *et al.*, 2018) melakukan klasifikasi pada empat klasifikasi risiko hipertensi. Klasifikasi risiko hipertensi tersebut adalah normal, prahipertensi, hipertensi derajat 1, dan hipertensi derajat 2. *Data set* yang digunakan pada penelitian tersebut berjumlah 125 data sekunder dengan 15 fitur yang berasal dari hasil rekam medis pasien. Metode NWKNN menggunakan *Cosine Similarity* memberikan hasil akurasi yang lebih baik dari pada metode NWKNN dengan menggunakan *Euclidean Distance* pada penelitian tersebut. Penelitian tersebut memiliki tingkat akurasi terbaik sebesar 88% dengan menggunakan nilai $K=10$, $E=4$, dan data latih sebanyak 100 data.

2.1.3 (Andriansyah *et al.*, 2018) dengan penelitian yang berjudul “Klasifikasi Risiko Hipertensi Menggunakan *Fuzzy Decision Tree Iterative Dichotomiser 3* (ID3)” melakukan penelitian untuk mendapatkan aturan yang akan digunakan untuk klasifikasi risiko hipertensi. Penelitian ini menggunakan 12 fitur seperti umur, tekanan darah (sistolik dan diastolik), lingkar perut, tinggi badan, berat badan, berat massa indeks, merokok, konsumsi gula, konsumsi garam, olahraga, dan konsumsi kafein. Penelitian tersebut menunjukkan bahwa aturan terbaik berasal dari 30 data latih pada 11 aturan dengan nilai FCT 70% dan LDT 3%. Akurasi terbaik sebesar 80% berasal dari 20 data uji dengan 30 data latih menggunakan nilai FCT dan LDT.

2.2 Uraian Landasan Teori

2.2.1 Hipertensi

Hipertensi adalah keadaan seseorang mengalami peningkatan tekanan darah di atas nilai normal yang dapat menyebabkan rasa sakit hingga kematian. Tekanan darah normal berada pada 120/80 mmHg (Herwati & Sartika, 2014). Batasan nilai tekanan darah dikenal dengan ketetapan JNC VII (*The Seventh Report of The Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure*) (Triyanto, 2014).

Tekanan darah adalah daya darah yang mendesak dinding-dinding pembuluh darah. Tekanan darah pada seseorang bervariasi secara alami. Tekanan ini sesuai pembuluh darah dan denyut jantung (Susilo & Wulandari, 2011). Tekanan paling tinggi ketika ventrikel berkontraksi (tekanan sistolik) dan paling rendah ketika ventrikel berelaksasi (tekanan diastolik). Tekanan darah sistolik adalah tekanan pembuluh darah ketika jantung berkontraksi dan memompa darah maju ke dalam pembuluh darah. Tekanan diastolik adalah tekanan di dalam pembuluh darah ketika jantung dalam kondisi istirahat (*relax*) setelah kontraksi (Triyanto, 2014).

Hipertensi disebut pembunuh gelap atau *the silent killer*. Hipertensi dengan secara tiba-tiba dapat melekat pada seseorang tanpa diketahui gejalanya terlebih dahulu (Yonata *et al.*, 2016). Gejala penyakit hipertensi sering dikenal sebagai gejala penyakit yang tidak perlu dikhawatirkan, seperti hidung berdarah (*mimisan*), sakit kepala, sulit bernapas, mudah lelah, penglihatan kabur, telinga berdering, jantung

berdebar-debar, wajah memerah, sering buang air kecil pada malam hari, dan dunia terasa berputar (*vertigo*) (Susilo & Wulandari, 2011).

Penderita hipertensi berasal dari segala umur dan kelas kelompok ekonomi. Penyebab hipertensi disebabkan oleh dua faktor, yaitu faktor keturunan dan pola hidup. Saat ini, penderita hipertensi yang disebabkan oleh faktor keturunan (genetika) belum diketahui letak kerusakan genetiknya. Penderita hipertensi oleh faktor pola hidup disebabkan karena banyak mengonsumsi garam, alkohol, merokok, obesitas, stres, dan diabetes (Rahajeng & Tuminah, 2009). Peningkatan dari tekanan darah sistolik dan diastolik dapat meningkatkan risiko kerusakan mata, penyakit jantung (*cardiac*), kerusakan otak (*stroke*), pengerasan pembuluh darah (*atherosclerosis*), penggumpalan pembuluh darah (*arteriosclerosis*), dan penyakit ginjal (*renal*) (Triyanto, 2014).

Pengobatan hipertensi bersifat jangka panjang dan seumur hidup. Pendeteksian penyakit hipertensi selama ini hanya dilakukan secara manual. Oleh karena itu, deteksi dini sangat diperlukan untuk meningkatkan perawatan kesehatan umum (Susilo & Wulandari, 2011). Tabel 1 adalah klasifikasi tekanan darah menurut (Triyanto, 2014) berdasarkan tekanan darah sistolik dan diastolik.

Tabel 1. Klasifikasi Tekanan Darah (Triyanto, 2014)

No.	Kategori	Sistolik (mmHg)	Diastolik (mmHg)
1	Normal	<130	<85
2	Normal Tinggi	130-139	85-89
3	Stadium 1 (Hipertensi Ringan)	140-159	90-99
4	Stadium 2 (Hipertensi Sedang)	160-179	100-109
5	Stadium 3 (Hipertensi Tinggi)	180-209	110-119
6	Stadium 4 (Hipertensi Maligna)	≥ 210	≥ 120

2.2.2 Machine Learning

Machine learning adalah metode yang mengajarkan mesin bekerja dengan *data set* (kumpulan data) secara lebih efisien (Dey, 2016). Penggunaan *machine learning* sangat tepat pada kerumitan variabel dan jumlah yang berbeda (Osisanwo *et al.*, 2017). *Machine learning* bekerja secara generalisasi, yaitu bekerja dengan membuat model yang mampu memprediksi data yang belum pernah dilihat (*unseen instances*) dengan tepat (Jan & Gotama, 2018).

Proses *machine learning* dibagi menjadi data *training* dan data *testing*. Data *training* (data pembelajaran) adalah himpunan data yang digunakan untuk membangun atau melatih model pembelajaran. Data *training* harus mengoptimalkan parameter. Data *testing* (data pengujian) adalah himpunan data yang digunakan untuk menguji model setelah proses latihan selesai (Mohammed *et al.*, 2016). Data *training* dan data *testing* tersusun dari data yang independen satu sama lain (tidak beririsan) dan memiliki karakteristik yang sama (*identically distributed*) untuk memastikan model yang dihasilkan memiliki generalisasi yang cukup baik (Jan & Gotama, 2018). *Machine learning* dibagi menjadi dua, yaitu:

- *Supervised Learning*

Supervised learning adalah algoritme yang membutuhkan bantuan eksternal yang memberikan pembelajaran (Raut *et al.*, 2017). *Supervised learning* memprediksi kelas berdasarkan *feature vector* yang berhubungan dengan kelas tertentu untuk mewakili suatu *instance* (Jan & Gotama, 2018). *Supervised learning* digunakan untuk klasifikasi atau prediksi (*regresi*). Klasifikasi (*classification*) adalah metode menemukan model untuk memprediksi kelas dari objek yang label kelasnya tidak diketahui

(Mase *et al.*, 2018). Prediksi (*regresi*) adalah metode yang digunakan oleh mesin untuk memprediksi suatu data berdasarkan data yang sudah dipelajari pada data *training*. Algoritme yang sering digunakan dalam *supervised learning* adalah *decision tree*, *naive bayes*, dan *support vector machine* (Mohammed *et al.*, 2016).

- *Unsupervised Learning*

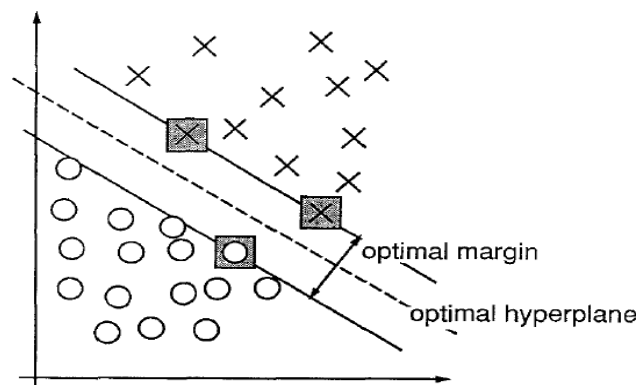
Unsupervised learning adalah algoritme untuk mempelajari beberapa fitur dari data. Fitur yang dipelajari sebelumnya digunakan untuk mengenali kelas data. *Unsupervised learning* digunakan untuk mengetahui variasi, pra-pemrosesan, dan struktur grup dari data yang tidak berlabel (Jan & Gotama, 2018; Sathya & Abraham, 2013). *Unsupervised learning* digunakan untuk pengelompokan (*clustering*) dan pengurangan fitur (*feature reduction*). Algoritme yang sering digunakan dalam pengelompokan dan pengurangan fitur adalah *k-means* dan *principal component analysis* (Dey, 2016).

2.2.3 Support Vector Machine (SVM)

Support vector machine (SVM) ditemukan oleh Vapnik pada tahun 1995. *Support vector machine* adalah metode untuk menemukan fungsi pemisah (*classifier*) yang optimal dari dua kelas yang berbeda. Penggunaan *support vector machine* akan menghasilkan proses komputasi yang cepat karena menggunakan *support vector* (Cortes & Vapnik, 1995). Berdasarkan hasil penelitian (Osisanwo *et al.*, 2017) dikemukakan bahwa *support vector machine* adalah algoritme presisi dan paling akurat pada klasifikasi. *Support vector machine* bekerja berdasarkan perhitungan margin dengan menarik margin di antara

kelas, sehingga jarak antara margin dan kelas maksimum dalam meminimalkan kesalahan (Dey, 2016).

Pada *support vector machine* penggunaan sampel berukuran besar diperlukan untuk mencapai akurasi maksimum tanpa memperhatikan jumlah atribut dan contoh data. Proses memaksimalkan margin dengan jarak terbesar antara pemisah *hyperplane* di kedua sisi akan mengurangi batas atas pada kesalahan generalisasi yang diharapkan (Osisanwo *et al.*, 2017). *Hyperplane* merupakan garis yang menjadi pemisah dari kelas satu dengan kelas lain. *Hyperplane* terbaik adalah *hyperplane* yang memiliki margin maksimal yang memiliki jarak terjauh terhadap data pada suatu kelas. *Hyperplane* yang memiliki margin besar dapat melakukan klasifikasi dengan baik pada data yang belum terlihat (*unseen instances*). Margin maksimal *classifier* bergantung pada *subset* data *training* yang dipilih sebagai *support vectors*. Maksimal margin *classifier* sangat sensitif terhadap perubahan data *training* (Jan & Gotama, 2018). Pada Gambar 2 ditunjukkan pemisah antar *class* oleh *hyperplane* untuk mencari margin yang optimal.



Gambar 2. Pendekatan konseptual pada metode *support vector machine* (Cortes & Vapnik, 1995).

Kernel berfungsi untuk menganalisis pola pada *support vector machine*. *Kernel* menghitung kemiripan antar vektor dalam dimensi transformasi (Bhavsar & Panchal, 2012). Terdapat tiga *kernel* dasar pada *support vector machine*, yaitu:

- **Linear**

Kernel linear adalah fungsi *kernel* yang paling sederhana. *Kernel* linear digunakan dalam permasalahan linear (Hossin & Sulaiman, 2015). *Kernel* linear memiliki nilai x , y dan konstanta opsional c . *Kernel* linear dihitung dengan Persamaan 1.

$$K(x_i, x_j) = x_i \cdot x_j + c \dots\dots\dots (1)$$

- **Polynomial**

Kernel polynomial adalah *kernel* yang tidak stasioner. *Kernel polynomial* digunakan untuk data pelatihan normalisasi (Hossin & Sulaiman, 2015). *Kernel polynomial* akan membuat model konjungsi fitur hingga urutan *polynomial* (Bhavsar & Panchal, 2012). Pada *kernel polynomial* nilai γ , r dan d adalah parameter *kernel*. *Kernel polynomial* dihitung dengan Persamaan 2.

$$K(x_i, x_j) = (\gamma x_i \cdot x_j + r)^d, \gamma > \dots\dots\dots (2)$$

- **Gaussian**

Kernel gaussian adalah *kernel* yang digunakan untuk proses yang tidak dapat diselesaikan pada *kernel* linear (Hossin & Sulaiman, 2015). Nilai γ , r dan d adalah parameter *kernel* yang dihitung secara *exponential function* dalam *kernel gaussian*. *Kernel gaussian* dihitung dengan Persamaan 3.

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2), \gamma > 0 \dots\dots\dots (3)$$

2.2.4 Feature Transformation

Feature transformation adalah proses menyederhanakan informasi dari sejumlah *data set* yang tidak berguna. Pemilihan *data set* yang paling berguna dari *data set* asli dapat membantu dalam proses klasifikasi. *Data set* yang telah dilakukan *feature transformasi* akan memiliki ukuran data yang lebih kecil. Proses *feature transformasi* sangat penting dalam penentuan *data set* terbaru dengan hasil berupa data variabel numerik (Liu & Motoda, 1998).

2.2.5 Analisis Korelasi Fitur

Analisis korelasi fitur adalah metode menentukan keterkaitan hubungan antara variabel satu dengan variabel yang lain. Pendekatan analisis korelasi fitur disebut koefisien korelasi. Koefisien korelasi berfungsi untuk mengetahui derajat hubungan antar variabel. Nilai koefisien korelasi berada diantara $-1 < 0 < 1$. Nilai korelasi -1 (negatif 1) adalah nilai korelasi negatif sempurna yang menjelaskan bahwa korelasi tersebut memiliki hubungan yang berlawanan antar variabel. Jika nilai koefisien korelasi bernilai 0 menjelaskan bahwa korelasi tidak memiliki hubungan antar variabel. Nilai korelasi +1 (positif 1) adalah nilai korelasi positif sempurna yang menjelaskan bahwa korelasi memiliki hubungan yang searah antar variabel (Vielhauer & Steinmetz, 2004).

Penelitian (Iin dan Sugiarto, 2012) pada Tabel 2 menunjukkan bahwa acuan koefisien korelasi untuk mengetahui hubungan korelasi antar variabel.

Tabel 2. Interpretasi Koefisien Korelasi (Iin & Sugiarto, 2012)

Besar Koefisien Korelasi (Positif atau Negatif)	Interpretasi Koefisien Korelasi
0,00	Tidak Ada Korelasi
0,01 – 0,20	Korelasi Sangat Lemah
0,21 – 0,40	Korelasi Lemah
0,41 – 0,70	Korelasi Sedang
0,71 – 0,99	Korelasi Tinggi
1,00	Korelasi Sempurna

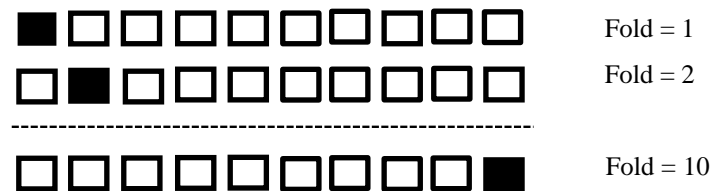
2.2.6 Cross Validation

Cross validation adalah *Cross validation* adalah metode statistik untuk mengevaluasi kinerja model yang memiliki generalisasi dan stabilitas yang baik pada data *training*. Model tersebut diuji kembali menggunakan data *testing* (Jan & Gotama, 2018). Beberapa *cross validation* yang sering digunakan adalah sebagai berikut:

- *K-Fold Cross Validation*

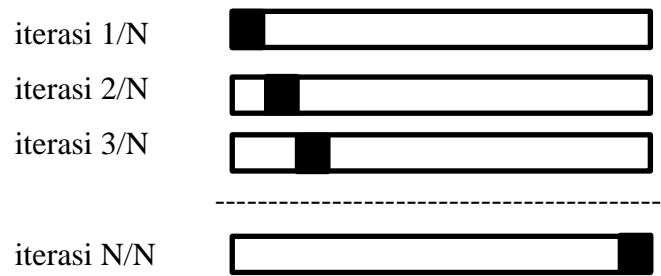
K-fold cross validation adalah teknik membagi *data set* menjadi k bagian dengan nilai $k=N$ (jumlah data). Pelatihan model menggunakan $k-1$ (k dikurang 1) bagian dapat menguji model pada satu bagian lainnya dalam data *testing*. Distribusi *subset* data harus berjumlah sama dengan *stratified sampling* agar menjadi beberapa *subset* partisi sebanyak k lipatan (*k-fold*). Stabilitas model diukur dengan menghitung variasi kinerja dari k model. Akurasi yang diperoleh dalam setiap iterasi (perulangan) akan dirata-ratakan untuk mendapatkan akurasi model. *10-fold cross validation* (k berjumlah 10) adalah *k-fold cross validation* yang sering digunakan pada *machine learning*. Model dengan kinerja

terbaik digunakan untuk proses data *testing* (Lukito & Chrismanto, 2015). Gambar 3 adalah *10-fold cross validation* dengan membagi sebanyak 10 lipatan (k berjumlah 10).



Gambar 3. Proses *k-fold cross validation* (Lukito & Chrismanto, 2015).

- *Stratified K-fold Cross Validation*
Stratified k-fold cross validation digunakan untuk variasi dalam teknik validasi silang. Setiap lipatan merupakan persentase sampel yang sama dari setiap kelas target. Validasi tidak menghitung semua cara untuk memecah sampel asli (Banjarsari *et al.*, 2016).
- *Leave-One-Out Cross Validation*
Leave-one-out cross validation (LOOCV) adalah jumlah *fold* sama dengan jumlah data *training*. Pada setiap iterasi (perulangan) digunakan untuk pelatihan dan modelnya diuji pada pengamatan tunggal. Tahap tersebut tidak dilakukan untuk observasi tunggal. *Leave-one-out cross validation* (LOOCV) digunakan untuk memperkirakan kemampuan generalisasi penggolongan statistik (Cawley & Talbot, 2003). Gambar 4 menunjukkan proses *leave-one-out cross validation* sejumlah n iterasi.



Gambar 4. Proses *leave-one-out cross validation* (Cawley & Talbot, 2003).

2.2.7 Confusion Matrix

Evaluasi matriks berfungsi untuk mendapatkan jumlah ketepatan klasifikasi terhadap kelas dengan algoritme yang dipakai (Sasongko, 2016). Menurut (Boughorbel *et al.*, 2017) pada Tabel 3, *confusion matrix* menunjukkan empat perkiraan dengan hasil yang berbeda, yaitu:

- TP (*True Positive*) adalah jumlah data positif yang terklasifikasi benar oleh sistem.
- FP (*False Positive*) adalah jumlah data positif yang terklasifikasi salah oleh sistem.
- FN (*False Negative*) adalah jumlah data negatif yang terklasifikasi salah oleh sistem.
- TN (*True Negative*) adalah jumlah data negatif yang terklasifikasi benar oleh sistem.

Tabel 3. Klasifikasi Evaluasi Matriks (Boughorbel *et al.*, 2017)

		<i>Actual Class</i>	
		<i>Positive</i>	<i>Negative</i>
<i>Predicted Class</i>	<i>Positive</i>	<i>True Positive (TP)</i>	<i>False Positive (FP)</i>
	<i>Negative</i>	<i>False Negative (FN)</i>	<i>True Negative (TN)</i>

Pada evaluasi matriks, tingkat *accuracy*, *presisi*, dan *recall/sensitivity* dipengaruhi oleh nilai TP (*True Positive*) dan TN (*True Negative*) (Jan & Gotama, 2018). Terdapat beberapa evaluasi matriks, yaitu:

- *Accuracy*

Accuracy adalah parameter yang menilai efektivitas rasio nilai benar atas keseluruhan data. *Accuracy* banyak digunakan dalam *biner classification* dan *multi-class classification*. Kelebihan *accuracy* adalah perhitungan kompleksitas yang lebih sedikit dan mudah dimengerti oleh manusia. *Accuracy* memiliki keterbatasan dalam membedakan pola dan proses evaluasi. Kelemahan tersebut menyebabkan kurang dapat memilih dan menentukan *classifier* optimal (Kusumodestoni & Sarwido, 2017; Hossin & Sulaiman, 2015). *Accuracy* dihitung dengan Persamaan 4.

$$Accuracy = (TP + TN) / (TP + FP + FN + TN) \times 100 \% \dots\dots\dots (4)$$

- *Recall* atau *Sensitivity*

Recall/sensitivity adalah parameter yang mengukur rasio prediksi benar positif dibandingkan dengan keseluruhan data yang benar positif (Kusumodestoni & Sarwido, 2017). *Recall/sensitivity* digunakan untuk menstabilkan dan mengoptimalkan dua kelas yang tidak seimbang (Hossin & Sulaiman, 2015). *Recall/sensitivity* tidak sensitif terhadap perubahan distribusi data (Bekkar *et al.*, 2013). *Recall/sensitivity* dapat dihitung dengan Persamaan 5.

$$Recall/sensitivity = (TP) / (TP + FN) \times 100 \% \dots\dots\dots (5)$$

- *Precision*

Precision adalah parameter untuk mengukur pola positif yang diprediksi dengan benar dari total pola yang diprediksi di kelas positif untuk menunjukkan tingkat ketelitian dalam pengklasifikasian (Sasongko, 2016). *Precision* tidak sensitif terhadap perubahan distribusi data (Bekkar *et al.*, 2013). *Precision* dapat dihitung dengan Persamaan 7.

$$Precision = (TP) / (TP + FP) \times 100 \% \dots\dots\dots (6)$$

III. METODOLOGI PENELITIAN

3.1 Tempat Dan Waktu Penelitian

3.1.1 Tempat

- **Tempat Pengumpulan Data**

Data pada penelitian ini berasal dari Dinas Kesehatan Kota Bandar Lampung yang beralamat di Jalan Way Pengubuan No.4, Pahoman, Kecamatan Teluk Betung Utara, Kota Bandar Lampung, Lampung 35228. Terlebih dahulu, dilakukan pencarian data dari lima puskesmas di Kota Bandar Lampung untuk divalidasi kebenaran data oleh Dinas Kesehatan Kota Bandar Lampung. Lima puskesmas untuk mencari data penyakit hipertensi adalah sebagai berikut:

- a. Puskesmas Simpur beralamat di Jalan Tamin No.121, Klp.Tiga, Kec. Tj. Karang Pusat, Kota Bandar Lampung, Lampung 35151.
- b. Puskesmas Way Halim beralamat di Jalan Gn. Rajabasa Raya No.3, Perumnas Way Halim, Way Halim, Kota Bandar Lampung, Lampung 35142.
- c. Puskesmas Kemiling beralamat di Jalan Beringin Raya, Kec. Kemiling, Kota Bandar Lampung, Lampung 35155.

- d. Puskesmas Pasar Ambon beralamat di Jalan Laksamana Malahayati No.11, Talang, Kec. Telukbetung Selatan, Kota Bandar Lampung, Lampung 35221.
 - e. Puskesmas Sukaraja beralamat di Jalan Yos Sudarso No.242, Sukaraja, Bumi Waras, Kota Bandar Lampung, Lampung 35226.
- Tempat Penelitian
Penelitian dilakukan di Laboratorium Rekayasa Perangkat Lunak (RPL) Jurusan Ilmu Komputer Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung yang beralamat di Jalan Prof. Dr. Soemantri Bojonegoro No.1 Gedung Meneng, Bandar Lampung.

3.1.2 Waktu

Penelitian dilaksanakan pada semester genap tahun ajaran 2020/2021, yaitu pada awal Maret 2020 hingga akhir bulan Maret 2021. Tahap pertama adalah pengumpulan data yang dilakukan pada bulan Maret sampai akhir April. Proses *cleaning* data dilakukan pada minggu ke-2 Juni sampai minggu ke-3 Juli. Data dilakukan proses *feature transformation* pada bulan Agustus. Selanjutnya, penggunaan *k-fold cross validation* pemodelan dilakukan pada minggu ke-1 September sampai minggu ke-1 Oktober. Tahap pemodelan dilakukan pada minggu ke-2 Oktober sampai akhir November. Terakhir, tahap evaluasi dilakukan pada bulan Desember sampai minggu ke-1 Februari 2021.

Pembuatan laporan penelitian dilakukan bersamaan dengan penelitian yaitu bulan Maret 2020 sampai Maret 2021. Studi literatur dan

penulisan bab I-III dilakukan pada bulan Maret sampai minggu ke-1 Juni. Seminar usul dilaksanakan pada akhir Juli. Pada minggu ke-2 Februari sampai minggu ke-3 Maret dilakukan untuk menyusun laporan bab IV-V. Terakhir, seminar hasil dilaksanakan pada minggu ke-4 Maret 2021. Penelitian ini selesai pada akhir Maret 2021. Jadwal penelitian dapat dilihat pada Tabel 4.

Tabel 4. *Gantt Chart* Waktu Pelaksanaan Penelitian Klasifikasi Hipertensi

Jadwal kegiatan	Bulan Pelaksanaan Penelitian 2020/2021																																																			
	Maret				April				Mei				Juni				Juli				Agustus				September				Oktober				November				Desember				Januari				Februari				Maret			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4				
1. Studi Literatur & Penulisan bab 1-3	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█																																					
2. Pengumpulan Data	█	█	█	█	█	█	█	█	█	█																																										
3. Cleaning Data											█	█	█	█	█	█																																				
4. Seminar Usul Penelitian																						█																														
5. Feature Transformation																						█	█	█	█	█																										
6. K-Fold Cross Validation																						█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█								
7. Membangun Model SVM Klasifikasi																																																				
8. Testing / Evaluasi Kinerja																																																				
9. Penulisan Laporan Bab 4-5																																																				
10. Seminar Hasil Penelitian																																																				

3.2 Alat Dan Data

3.2.1 Alat

- Perangkat Keras
 - a. Prosesor Intel® Core™ i3-7020U CPU @2.30GHz 2.30GHz.
 - b. Tipe sistem 64-bit operating system, x64-based Processor.

- c. VGA NVidia GeForce MX230 2GB.
 - d. 4 GB DDR4.
- Perangkat Lunak
 - a. Sistem operasi *Windows 10 Pro N 64-bit (10,0, Build 17134)* untuk mengatur berjalannya perangkat keras pada sistem komputer.
 - b. *Microsoft Excel 2016* untuk menyimpan *data set*.
 - c. *R for Windows* versi 3.5.1 sebagai bahasa pemrograman.
 - d. *R studio* versi 3.5.1 sebagai perangkat lunak pengolahan data dengan menggunakan bahasa pemrograman R.
 - e. *Package caret* untuk *confusion matrix*.
 - f. *Package e1701* untuk mendukung instalasi *package caret*.
 - g. *Package hmisc* untuk menampilkan korelasi fitur.
 - h. *Package readr* untuk *import data*.

3.2.2 Data

Data penelitian klasifikasi hipertensi di Kota Bandar Lampung adalah data penyakit hipertensi di Kota Bandar Lampung tahun 2017 sampai dengan tahun 2019. Data penyakit hipertensi di Kota Bandar Lampung yang digunakan pada penelitian ini berjumlah 919 data. Data tersebut terdiri dari 609 data negatif hipertensi dan 310 data positif hipertensi.

Data penyakit hipertensi dari lima puskesmas di Kota Bandar Lampung terdiri dari 11 atribut. Pada Tabel 5 dikemukakan bahwa data penyakit hipertensi meliputi data numerik dan data kategorik. Data numerik penyakit hipertensi adalah usia, sistolik, diastolik, tinggi badan, berat badan, dan lingkar perut. Data kategorik penyakit

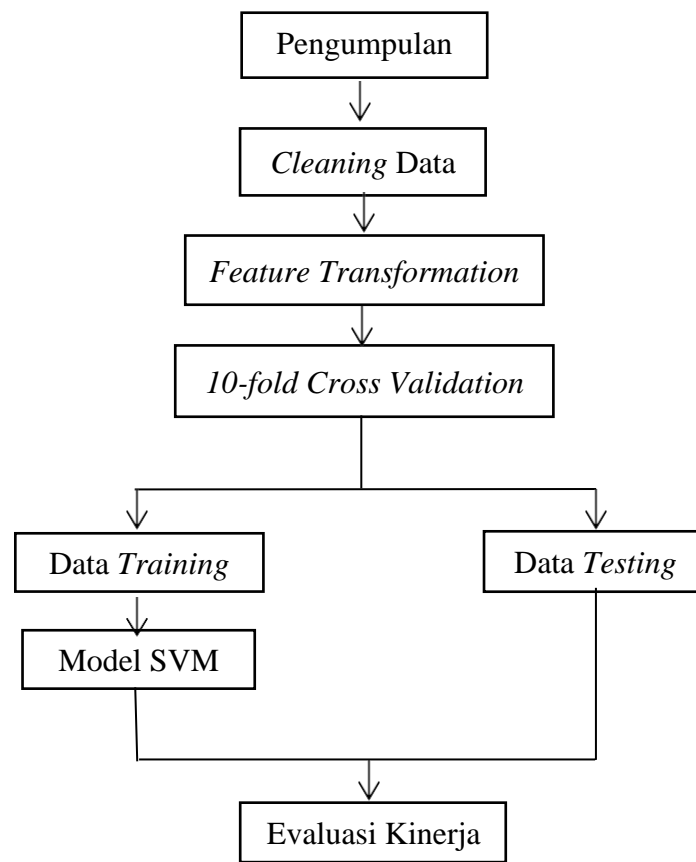
hipertensi adalah merokok, kurang aktivitas fisik, kurang makan sayur buah, konsumsi alkohol, dan obesitas.

Tabel 5. *Data Set* Penelitian Klasifikasi Hipertensi

No.	Atribut	Tipe Data
1	Usia	Numerik, (18-68; 38.19)
2	Sistolik	Numerik, (90-190; 124.31), mmHg
3	Diastolik	Numerik, (60-130; 78.92), mmHg
4	Tinggi Badan	Numerik, (145-175; 159.13), centimeter
5	Berat Badan	Numerik, (47-87; 62.87), centimeter
6	Lingkar Perut	Numerik, (71-111; 84.63), centimeter
7	Merokok	Kategorik, (Ya, Tidak)
8	Obesitas	Kategorik, (Ya, Tidak)
9	Kurang Makan Sayur Buah	Kategorik, (Ya, Tidak)
10	Konsumsi Alkohol	Kategorik, (Ya, Tidak)
11	Kurang Aktivitas Fisik	Kategorik, (Ya, Tidak)

3.3 Metode

Penelitian klasifikasi penyakit hipertensi di Kota Bandar Lampung dengan metode *support vector machine* (SVM) terdiri dari beberapa tahap. Tahapan tersebut dijelaskan pada proses yang ditunjukkan pada Gambar 5.



Gambar 5. *Flowchart* tahap penelitian klasifikasi hipertensi.

Berdasarkan proses klasifikasi pada Gambar 5, berikut ini adalah penjelasan dari setiap proses yang akan dilakukan pada penelitian, yaitu:

1. Pengumpulan Data

Data yang diperlukan pada penelitian adalah data penyakit hipertensi di Kota Bandar Lampung tahun 2017 sampai dengan tahun 2019. Data penyakit hipertensi di Kota Bandar Lampung yang digunakan pada penelitian berjumlah 919 data. Data tersebut terdiri dari 609 data negatif hipertensi dan 310 data positif hipertensi. Data yang terkumpul dari lima puskesmas sebelum dilakukan proses *cleaning* data dengan data yang terkumpul sebanyak 1003 data.

2. *Cleaning Data*

Proses *cleaning* data adalah tahap pengurangan atau penghapusan data yang tidak sesuai. Data tersebut adalah data penyakit hipertensi yang kosong, sehingga lebih mudah dalam pengolahan data klasifikasi penyakit hipertensi. Pada penelitian ini dihapus sebanyak 84 data hipertensi yang kosong dari total 1003 data hipertensi yang dikumpulkan dari lima puskesmas di Kota Bandar Lampung.

3. *Feature Transformation*

Pada proses ini dilakukan proses *feature transformation* untuk mengubah data nominal menjadi data numerik. Proses *feature transformation* dilakukan secara manual dengan memberi nilai pada dua kategori yaitu nilai Ya=1 dan nilai Tidak=0. Pada penelitian ini atribut merokok, kurang konsumsi sayur dan buah, kurang aktivitas, konsumsi alkohol dan obesitas dilakukan proses *feature transformation*.

4. *K-Fold Cross Validation*

Pada tahap ini dilakukan proses data *training* dan data *testing*. Pada uji validitas dilakukan *10-fold cross validation* dengan membagi sebanyak 10 lipatan. Pada data *training*, dilakukan proses *k-fold cross validation* sebesar 90% dari data. Data *testing* menggunakan sebesar 10% dari data yang tersedia. Pada penelitian ini, 827 data digunakan untuk proses data *training* dan 92 data digunakan untuk proses data *testing*. Proses akan diulangi sebanyak k lipatan (k berjumlah 10) dan rata-rata nilai akan diproses untuk ke tahap selanjutnya.

5. Model SVM

Tahap yang dilakukan pada proses ini adalah menentukan model dengan memilih tipe *kernel* yang tepat. Nilai yang terpilih akan dilakukan proses klasifikasi dengan menggunakan metode *support*

vector machine. Pada penelitian ini, digunakan *kernel* linear, *polynomial*, dan *gaussian*.

6. *Testing* / Pengujian

Tahap terakhir pada penelitian ini adalah melakukan evaluasi kinerja. Model *support vector machine* yang tersedia pada tahap sebelumnya akan dilakukan proses evaluasi. Evaluasi kinerja akan mendapatkan nilai tipe *kernel* dan nilai parameter yang baik pada klasifikasi. Pengukuran parameter pada penelitian ini dilakukan dengan menggunakan *confusion matrix* yang terdiri dari parameter *accuracy*, *sensitivity/recall*, dan *precision*. Kemudian, dilakukan perhitungan rata-rata *confusion matrix* untuk mendapatkan hasil kinerja klasifikasi kejadian hipertensi dengan metode *support vector machine* (SVM) menggunakan data puskesmas di Kota Bandar Lampung.

V. SIMPULAN DAN SARAN

5.1 Simpulan

Pada hasil penelitian ini, dapat menghasilkan simpulan sebagai berikut:

1. Penelitian memperoleh hasil klasifikasi dengan menggunakan metode *support vector machine* (SVM) dengan teknik penentuan parameter terbaik menggunakan *10-fold cross validation*. Proses pembagian data dilakukan dengan aturan 90% data *training* dan 10% data *testing* dari data yang tersedia. Terdapat 919 data pada penelitian ini dengan pembagian data *training* sebanyak 827 data dan data *testing* sebanyak 92 data.
2. Hasil klasifikasi hipertensi menggunakan metode *support vector machine* (SVM) didapatkan hasil rata-rata rata-rata *accuracy* pada *kernel linear* sebesar 99.67%, dengan hasil rata-rata *recall* sebesar 99.83%, dan hasil rata-rata *precision* sebesar 99.67%. Hasil rata-rata *accuracy* pada *kernel gaussian* sebesar 99.78%, dengan hasil rata-rata *recall* sebesar 99.83%, dan hasil rata-rata *precision* sebesar 99.83%. Hasil rata-rata *accuracy* pada *kernel polynomial* sebesar 99.30%, dengan hasil rata-rata *recall* sebesar 99.83%, dan hasil rata-rata *precision* sebesar 99.20%.
3. Keseluruhan percobaan hasil klasifikasi menggunakan tiga *kernel* didapatkan *accuracy* terbesar pada *kernel gaussian* yaitu sebesar 99.78%, *recall* tidak memiliki nilai terbesar pada ketiga *kernel* dikarenakan memiliki nilai yang sama yaitu sebesar 99.83%, dan *precision* terbesar pada *kernel gaussian* yaitu sebesar 99.83%.

5.2 Saran

Saran yang diberikan dalam penelitian ini adalah penelitian selanjutnya dapat menggunakan metode klasifikasi yang lain, agar dapat dibandingkan dengan hasil nilai akurasi metode SVM.

DAFTAR PUSTAKA

- Andriansyah, M. R., Santoso, E., & Sutrisno. 2018. Klasifikasi Risiko Hipertensi Menggunakan Fuzzy Decision Tree Iterative Dichotomiser 3 (ID3). *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 2(Vol 2 No 12 (2018)), 7088–7096.
- Arda, Z. A., Ali, R., & Mustapa, M. 2018. Hipertensi dan Faktor Risikonya di Puskesmas Motolohu Kabupaten Pohuwato. *Gorontalo Journal of Public Health*, 1(1), 032. <https://doi.org/10.32662/gjph.v1i1.148>.
- Banjarsari, M. A., Budiman, I., & Farmadi, A. 2016. Penerapan K-Optimal pada Algoritma KNN untuk Prediksi Kelulusan Tepat Waktu Mahasiswa Program Studi Ilmu Komputer Fmipa Unlam Berdasarkan IP sampai dengan Semester 4. *Klik - Kumpulan Jurnal Ilmu Komputer*, 2(2), 159–173. <https://doi.org/10.20527/KLIK.V2I2.26>.
- Bekkar, M., Djemaa, H. K., & Alitouche, T. A. 2013. Evaluation Measures for Models Assessment Over Imbalanced Data Sets. *Journal of Information Engineering and Applications*, 3(10), 27–38.
- Bhavsar, H., & Panchal, M. H. 2012. A Review on Support Vector Machine for Data Classification. *International Journal of Advanced Research in Computer Engineering & Technology*, 1(10), 2278–1323.
- Boughorbel, S., Jarray, F., & El-Anbari, M. 2017. Optimal Classifier for Imbalanced Data using Matthews Correlation Coefficient Metric. *PLoS ONE*, 12(6), 1–17. <https://doi.org/10.1371/journal.pone.0177678>.
- Cawley, G. C., & Talbot, N. L. C. 2003. Efficient Leave-one-out Cross Validation of Kernel Fisher Discriminant Classifiers. *Pattern Recognition*, 36(11), 2585–2592. [https://doi.org/10.1016/S0031-3203\(03\)00136-5](https://doi.org/10.1016/S0031-3203(03)00136-5).

- Cortes, C., & Vapnik, V. 1995. Photonic Neural Networks and Learning Machines. *IEEE Expert*, 7(5), 63–72. <https://doi.org/10.1109/64.163674>.
- Dey, A. 2016. Machine Learning Algorithms: A Review. *International Journal of Computer Science and Information Technologies*, 7(3), 1174–1179.
- Osisanwo, F. Y., Akinsola, J. E. T., Awodele, O., Hinmikaiye, J. O., Olakanmi, O., & Akinjobi, J. 2017. Supervised Machine Learning Algorithms: Classification and Comparison. *International Journal of Computer Trends and Technology*, 48(3), 128–138. <https://doi.org/10.14445/22312803/ijctt-v48p126>.
- Herwati, & Sartika, W. 2014. Terkontrolnya Tekanan Darah Penderita Hipertensi Berdasarkan Pola Diet dan Kebiasaan Olahraga di Padang Tahun 2011. *Jurnal Kesehatan Masyarakat*, 8(1), 9.
- Iin, Y., & Sugiarto, B. 2012. Korelasi Antara Keterampilan Metakognitif Dengan Hasil Belajar Siswa Di Sman 1 Dawarblandong, Mojokerto. *Unesa Journal Of Chemical Education*, 1, 78–83.
- Jan, P., & Gotama, W. 2018. *Pengenalan Pembelajaran Mesin dan Deep Learning Jan Wira Gotama Putra Pengenalan Konsep Pembelajaran Mesin dan Deep Learning*, 7(1), 1–199.
- Kementerian Kesehatan Republik Indonesia. 2018. Hasil Utama Riset Kesehatan Dasar. *Kementerian Kesehatan Republik Indonesia*, 1(1), 1–100.
- Kusumodestoni, R. H., & Sarwido, S. 2017. Komparasi Model Support Vector Machines (SVM) dan Neural Network untuk Mengetahui Tingkat Akurasi Prediksi Tertinggi Harga Saham. *Jurnal Informatika Upgris*, 3(1). <https://doi.org/10.26877/jiu.v3i1.1536>.
- Liu, H., & Motoda, H. 1998. Feature Transformation and Subset Selection. *IEEE Expert*, 13(2), 26–28. <https://doi.org/10.1109/mis.1998.671088>.
- Lukito, Y., & Chrismanto, A. R. 2015. Perbandingan Metode-Metode Klasifikasi untuk Indoor Positioning System. *Jurnal Teknik Informatika dan Sistem Informasi*, 1(2), 123–131. <https://doi.org/10.28932/jutisi.v1i2.373>.
- Hossin, M., & Sulaiman, M. N. 2015. A Review on Evaluation Metrics for Data Classification Evaluations. *International Journal of Data Mining &*

Knowledge Management Process, 5(2), 01-11.
<https://doi.org/10.5121/ijdkp.2015.5201>. Hossin, M., & Sulaiman, M. N.

Mase, J., Furqon, M. T., & Rahayudi, B. 2018. Penerapan Algoritme Support Vector Machine (SVM) pada Pengklasifikasian Penyakit Kucing. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer (J-PTIIK) Universitas Brawijaya*, 2(10), 3648–3654.

Mohammed, M., Khan, M. B., & Bashier, E. B. M. 2016. *Algorithms and Applications*. <https://doi.org/10.1201/9781315371658>.

Rahajeng, E., & Tuminah, S. 2009. Hidup Bersama Hipertensi. *Maj Kedokteran Indonesia*, 59(12), 580–587.

Raut, P. P., Borkar & Borkar, N. R. 2017. Machine Learning Algorithms:Trends, Perspectives and Prospects. *International Journal of Engineering Science and Computing*, 7(3), 4884–4891.

Samant, R., & Rao, S. 2013. A Study on Comparative Performance of SVM Classifier Models with Kernel Functions in Prediction of Hypertension. *International Journal of Computer Science and Information Technologies*, 4(6), 818–821.

Sasongko, T. B. 2016. Komparasi dan Analisis Kinerja Model Algoritma SVM dan PSO-SVM (Studi Kasus Klasifikasi Jalur Minat SMA). *Jurnal Teknik Informatika dan Sistem Informasi*, 2(2), 244-253.
<https://doi.org/10.28932/jutisi.v2i2.476>.

Sathya, R., & Abraham, A. 2013. *Comparison of Supervised and Unsupervised Learning Algorithms for Pattern Classification*, 2(2), 34–38.

Susilo, Y., & Wulandari, A. 2011. *Cara Jitu Mengatasi Hipertensi*. Yogyakarta:C.V Andi Offset

Triyanto, E. 2014. *Pelayanan Keperawatan bagi Penderita Hipertensi secara Terpadu*. Yogyakarta:Graha Ilmu.

Vielhauer, C., & Steinmetz, R. 2004. Handwriting : Feature Correlation Analysis for Biometric Hashes. *Eurasip Journal on Applied Signal Processing*, 4, 542- 558.

- Yonata, A., Satria, A., & Pratama, P. 2016. Hipertensi sebagai Faktor Pencetus Terjadinya Stroke, 5(3), 17.
- Yudha, B. L., Muflikhah, L., & Wihandika, R. C. 2018. Klasifikasi Risiko Hipertensi Menggunakan Metode Neighbor Weighted K-Nearest Neighbor (NWKNN). *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer (J-PTIIK) Universitas Brawijaya*, 2(2), 897–904.