

**KLASIFIKASI DIARE NOSOKOMIAL PADA LANSIA  
MENGUNAKAN METODE *SUPPORT VECTOR MACHINE***

**(Skripsi)**

**Oleh**

**Yolanda Meilani Sirait**



**JURUSAN ILMU KOMPUTER  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS LAMPUNG  
2021**

## **ABSTRAK**

### **KLASIFIKASI DIARE NOSOKOMIAL PADA LANSIA MENGGUNAKAN SUPPORT VECTOR MACHINE**

**Oleh**

**YOLANDA MEILANI SIRAIT**

Diare merupakan penyakit potensial Kejadian Luar Biasa (KLB) yang sering disertai dengan kematian. Satu faktor infeksi penyebab diare adalah infeksi nosokomial, yaitu infeksi yang didapat dari rumah sakit. Beberapa faktor risiko infeksi diare nosokomial adalah usia, lama rawat inap, jumlah antibiotik, dan pemberian pipa/selang makanan atau pipa nasogastrik, paparan kemoterapi sitotoksik, penggunaan terapi penekan asam, penghambat pompa proton (PPIs), kontak dengan sistem perawatan kesehatan. Penelitian ini bertujuan untuk mengetahui hubungan antara faktor risiko dan diare nosokomial dengan klasifikasi menggunakan algoritma Support Vector Machine. Penelitian ini menggunakan 218 dataset lansia dengan 48 diare dan 170 non-diare. Penelitian ini mengimplementasikan metode Support Vector Machine dengan 10-Fold Cross Validation dan package boruta sebagai feature selection. Klasifikasi pada penelitian ini mendapatkan AUC terbaik oleh kernel Gaussian, dengan 93.81% data train dan 83% data tes untuk percobaan tanpa feature selection, dan 95.30% data train dan 85.91% data tes untuk percobaan dengan feature selection.

**Kata Kunci:** Klasifikasi, Diare Nosokomial, Algoritma Boruta, Support Vector Machine, AUC

## **ABSTRACT**

### **CLASSIFICATION OF NOSOCOMIAL DIARRHEA IN THE ELDERLY USING SUPPORT VECTOR MACHINE**

**By**

**YOLANDA MEILANI SIRAIT**

Diarrhea includes as an outbreak disease that often caused death. One significant effect of diarrhea infection is a nosocomial infection, which is an infection obtained from a hospital. Some of the risk factors for nosocomial diarrheal infections are age, length of hospitalization, several antibiotics, feeding tube/tube or nasogastric tube, exposure to cytotoxic chemotherapy, acid-suppressing therapy, proton pump inhibitors (PPIs), contact with the health care system. This study aims to determine the relationship between risk factors and nosocomial diarrhea with classification using the Support Vector Machine algorithm. This study used 218 elderly datasets with 48 diarrhea and 170 non-diarrhea. This research implements the Support Vector Machine method with 10-Fold Cross-Validation and packages boruta as feature selection. The classification in this study obtained the best AUC by the Gaussian kernel, with 93.81% train data and 83% test data for experiments without feature selection, and 95.30% train data and 85.91% test data for experiments with feature selection.

**Keywords:** Classification, Nosocomial Diarrhea, Boruta Algorithm, Support Vector Machine, AUC

**KLASIFIKASI DIARE NOSOKOMIAL PADA LANSIA MENGGUNAKAN  
SUPPORT VECTOR MACHINE**

**Oleh**

**YOLANDA MEILANI SIRAIT**

**Skripsi**

**Sebagai Salah Satu Syarat untuk Memperoleh Gelar  
SARJANA KOMPUTER**



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

Judul Skripsi

: **KLASIFIKASI DIARE NOSOKOMIAL PADA  
LANSIA MENGGUNAKAN METODE  
SUPPORT VECTOR MACHINE**

Nama Mahasiswa

: **Yolanda Meilani Sirait**

Nomor Pokok Mahasiswa

: 1617051127

Program Studi


: Ilmu Komputer

Fakultas

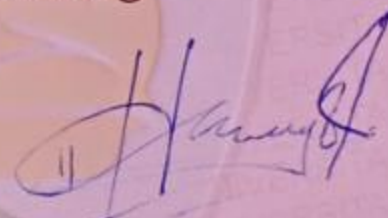
: Matematika dan Ilmu Pengetahuan Alam

**MENYETUJUI**

1. Komisi Pembimbing



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



**Yunda Heningtyas, S.Kom., M.Kom.**  
NIP 19890108 201903 2 014

2. Ketua Jurusan Ilmu Komputer



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

## MENGESAHKAN

### 1. Tim Penguji

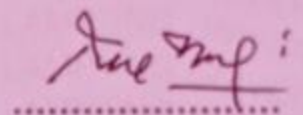
Ketua : **Favorisen R. Lumbanraja, Ph.D.**



Sekretaris : **Yunda Heningtyas, S.Kom., M.Kom.**



Penguji  
Bukan Pembimbing : **Dr. Ir. Kurnia Muludi, M.S.Sc.**



### 2. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam



**Dr. Eng. Supto Dwi Yuwono, S.Si., M.T.**

NIP 19740705 200003 1 001

Tanggal Lulus Ujian Skripsi : **02 Agustus 2021**

## PERNYATAAN

Saya yang bertanda tangan di bawah ini, menyatakan bahwa skripsi saya dengan judul "Klasifikasi Diare Nosokomial Pada Lansia Menggunakan Support Vector Machine" merupakan karya saya sendiri dan bukan karya orang lain. Semua tulisan yang tertuang dalam skripsi ini telah mengikuti kaidah penulisan karya ilmiah Universitas Lampung. Apabila dikemudian hari terbukti skripsi saya merupakan hasil penjiplakan atau dibuat orang lain, maka bersedia menerima sanksi berupa pencabutan gelar yang telah saya terima.

Bandar Lampung, 30 Agustus 2021



Yolanda Meilani Sirait

NPM. 1617051127

## **RIWAYAT HIDUP**



Penulis dilahirkan pada tanggal 16 Mei 1999 di Jakarta, sebagai anak kedua dari tiga bersaudara dengan Ayah bernama Liser Sirait dan Ibu bernama Ronika Nadeak.

Penulis menyelesaikan pendidikan Sekolah Dasar di SD Negeri Parapat 2 Kota Tangerang pada tahun 2010. Penulis melanjutkan pendidikan menengah pertama di SMP Negeri 19 Kota Tangerang dan selesai pada tahun 2013. Kemudian penulis melanjutkan pendidikan menengah atas di SMA Negeri 5 Kota Tangerang dan selesai pada tahun 2016.

Pada tahun 2016, penulis terdaftar sebagai mahasiswa Jurusan Ilmu Komputer Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung melalui jalur SBMPTN. Selama masa perkuliahan, penulis mengikuti organisasi internal jurusan yaitu Himakom (Himpunan Mahasiswa Jurusan Ilmu Komputer) pada periode 2017/2018. Selama menjadi mahasiswa, beberapa kegiatan yang dilakukan penulis antara lain pada bulan Januari 2017 penulis melaksanakan Karya Wisata Ilmiah (KWI) di Desa Margosari Pagelaran Utara Pringsewu, lalu pada bulan Januari 2019 penulis melaksanakan Kerja Praktik di Kominfo Jakarta Utara, dan pada bulan Juli 2019 penulis melaksanakan Kuliah Kerja Nyata (KKN) di Desa Jabung, Kabupaten Lampung Timur.



## **PERSEMBAHAN**

Puji Syukur atas berkat dan rahmat Tuhan Yesus Kristus, penulis dapat menyelesaikan skripsi ini. Kupersembahkan Skripsi ini untuk orang-orang yang yang mencintai dan menyayangiku.

Terima kasih untuk orang tua yang senantiasa ada untuk penulis, terima kasih untuk tak pernah berhenti memberi nasihat, perhatian, semangat, motivasi dan doa sehingga skripsi ini dapat selesai. Serta untuk abang dan idung yang selalu mendukung memberikan semangat sehingga skripsi ini dapat diselesaikan.

Terima kasih kepada bapak dan ibu dosen Jurusan Ilmu Komputer yang telah membagikan ilmu dan motivasi selama perkuliahan.

Terima kasih teman-teman Ilmu Komputer 2016 yang selalu mendukung dan berbagi kebahagiaan dan kesedihan bersama, serta Almamater tercinta.

**UNIVERSITAS LAMPUNG**

## **MOTTO**

Mintalah, maka akan diberikan kepadamu.

Carilah, maka kamu akan mendapat.

Ketoklah, maka pintu akan dibukakan bagimu. (Matthew 7:7)

Karena masa depan sungguh ada,  
dan harapanmu tidak akan hilang. (Proverbs 23:18)

Yakin dan percaya. –mypapa

## UCAPAN TERIMA KASIH

Shalom dan Salam Sejahtera.

Puji syukur kepada Tuhan yang Maha Esa. Berkat kasih setia-Nya dan anugerah-Nya, penulis dapat menyelesaikan skripsi ini yang berjudul “Klasifikasi Diare Nosokomial Pada Lansia Menggunakan Support Vector Machine”. Skripsi ini merupakan satu syarat untuk memperoleh gelar Sarjana Komputer di Jurusan Ilmu Komputer Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung. Penulis sangat berterima kasih dan memberikan penghargaan yang sedalam-dalamnya kepada seluruh pihak yang membantu penulis menyelesaikan skripsi ini. Penulis ingin mengucapkan terima kasih dengan setulus hati terutama kepada:

1. Family; Mama, Bapa, Abang Doni, dan idung. Terima kasih sudah selalu ada.
2. Bapak Favorisen R. Lumbanraja, Ph.D., sebagai dosen pembimbing skripsi atas kesediaan, kesabaran, dan keikhlasan untuk memberikan dukungan, bimbingan, nasihat, saran, dan kritik dalam proses penyelesaian skripsi ini.
3. Ibu Yunda, S.Kom., M.Kom., sebagai Pembimbing Kedua yang telah memberikan bimbingan, saran, dan masukan yang sangat membantu dalam penyusunan dan perbaikan skripsi ini.
4. Bapak Dr. Ir. Kurnia Muludi, M.S.Sc., sebagai Pembahas yang telah memberikan masukan, bimbingan, dan nasihat dalam penyelesaian skripsi ini.
5. Bapak Bambang Hermanto, S.Kom., M.Cs., sebagai Pembimbing Akademik yang banyak memberikan bimbingan dan motivasi selama perkuliahan.
6. Bapak Didik Kurniawan, S.Si., M.T., selaku ketua Jurusan Ilmu Komputer FMIPA Universitas Lampung.
7. Bapak Dr. Eng. Suropto Dwi Yuwono, M.T., selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.

8. Bapak dan Ibu Dosen Jurusan Ilmu Komputer yang telah memberikan ilmu yang bermanfaat selama penulis menjadi mahasiswa.
9. Ibu Ade Nora Maela, Mas Naufal, Mas Irsan, Mas Zai dan Mbak Lia yang telah membantu memudahkan segala administrasi penulis di Jurusan Ilmu Komputer.
10. HIMAKOM, terutama teman-teman jurusan ilmu komputer 2016.
11. Universitas Lampung, yang telah memberikan penulis kesempatan untuk mendapatkan pendidikan perkuliahan S1.
12. Audhi Aprilliant, S.Stat. Terima kasih banyak banyak banyak, Audhi ^^
13. Devy Refina J. S.T. dan Angie Pricilia S.Ak., fellas since elementary school. 17 years and still counting. Terima kasih atas semangat dan dukungan.
14. Juniva Rahmi, S.Ak. 11 years still counting, peh^^
15. Tobias Danang, S.Kom., Tetra Dilla, S.Kom., Hartsa Hanifah, S.Kom. Nice to know you buddies.
16. Krist Perawat Sangpotirat, my treasure. I can get through bad things because of you.
17. Tay Tawan Vihokratana, my lowkey.
18. Singto Prachaya Ruangroj, my lion.
19. Wave Khoo, my humble, down-to-earth man. Thank, P'Wave.
20. Ohm Pawat, my nong.
21. Bank Thiti, Off Jumpol, Mike Chinnarat, Boun Noppanut, Billkin Putthipong, Gulf Kanawut, Mew Suppasit, Bright Vachirawit, Win Metawin, and they are all my Thai men.

Penulis menyadari skripsi ini masih jauh dari sempurna. Penulis berharap skripsi ini dapat memberikan manfaat, baik kepada penulis khususnya kepada pembaca.

Bandar Lampung, 30 Agustus 2021

Yolanda Meilani Sirait

NPM. 1617051127

## DAFTAR ISI

	<b>Halaman</b>
<b>DAFTAR ISI</b> .....	<b>xii</b>
<b>DAFTAR TABEL</b> .....	<b>xiv</b>
<b>DAFTAR GAMBAR</b> .....	<b>xv</b>
<b>DAFTAR KODE</b> .....	<b>xvi</b>
<b>I. PENDAHULUAN</b> .....	<b>1</b>
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah.....	3
1.3 Tujuan .....	3
1.4 Batasan Masalah .....	3
1.5 Manfaat .....	4
<b>II. TINJAUAN PUSTAKA</b> .....	<b>5</b>
2.1 Penelitian Terdahulu .....	5
2.2 Diare.....	7
2.3 Infeksi Nosokomial.....	7
2.4 <i>Machine Learning</i> .....	8
2.5 <i>Feature Selection</i> .....	10
2.6 <i>Cross-Validation</i> .....	11
2.7 <i>Support Vector Machine</i> .....	12
2.8 <i>Confusion Matrix</i> .....	13
<b>III. METODE PENELITIAN</b> .....	<b>16</b>
3.1 Waktu dan Tempat Penelitian.....	16
3.2 Data dan Alat .....	17
3.3 Alur Kerja Penelitian .....	18
<b>IV. HASIL DAN PEMBAHASAN</b> .....	<b>23</b>

4.1 Percobaan tanpa <i>Feature Selection</i> .....	23
4.2 Percobaan dengan <i>Feature Selection</i> .....	29
4.3 Pembahasan.....	38
<b>V. SIMPULAN DAN SARAN</b> .....	<b>41</b>
5.1 Simpulan.....	41
5.2 Saran.....	41
<b>DAFTAR PUSTAKA</b> .....	<b>43</b>

## DAFTAR TABEL

<b>Tabel</b>	<b>Halaman</b>
1. Hasil Penelitian Terdahulu .....	7
2. <i>Confusion Matrix</i> .....	13
3. Kriteria Klasifikasi AUC .....	15
4. <i>Timeline</i> Waktu Penelitian .....	16
5. Hasil Kernel Linear tanpa <i>Feature Selection</i> .....	27
6. Hasil Kernel Polynomial tanpa <i>Feature Selection</i> .....	28
7. Hasil Kernel Gaussian tanpa <i>Feature Selection</i> .....	28
8. Hasil Percobaan Data Tes tanpa <i>Feature Selection</i> .....	29
9. Hasil Percobaan Kernel Linear dengan <i>Feature Selection</i> .....	36
10. Hasil Percobaan Kernel Polynomial dengan <i>Feature Selection</i> .....	36
11. Hasil Percobaan Kernel Gaussian dengan <i>Feature Selection</i> .....	37
12. Hasil Percobaan Data Tes dengan <i>Feature Selection</i> .....	37
13. Perbandingan Kinerja Klasifikasi .....	40

## DAFTAR GAMBAR

<b>Gambar</b>	<b>Halaman</b>
1. Penerapan Kriteria Eksklusi pada Penelitian Terdahulu .....	v
2. Machine Learning .....	9
3. Konsep Metode SVM .....	12
4. Grafik ROC .....	15
5. Alur Kerja Penelitian .....	19
6. Cara Kerja Algoritma Boruta .....	20
7. Ilustrasi Contoh Dataset .....	21
8. Hasil Feature Selection .....	33
9. Kinerja Klasifikasi SVM Linear, Polynomial, Gaussian .....	38
10. Perbandingan AUC dengan Feature Selection dan Tanpa Feature Selection .....	39



## DAFTAR KODE

<b>Kode</b>	<b>Halaman</b>
1. <i>Loading Package</i> .....	23
2. <i>Import Data</i> .....	24
3. <i>Pembagian Data</i> .....	24
4. <i>K-Fold Cross Validation</i> .....	25
5. <i>Membuat Model SVM</i> .....	26
6. <i>Kinerja Klasifikasi</i> .....	27
7. <i>Loading Package</i> .....	30
8. <i>Import Data</i> .....	30
9. <i>Pembagian Data</i> .....	30
10. <i>K-Fold Cross Validation</i> .....	31
11. <i>Feature Selection</i> .....	32
12. <i>Membuat Model SVM</i> .....	34
13. <i>Proses Klasifikasi</i> .....	35

# I. PENDAHULUAN

## 1.1 Latar Belakang

Diare merupakan penyakit potensial Kejadian Luar Biasa (KLB) yang sering disertai dengan kematian (Kementerian Kesehatan RI, 2019). Penyakit diare dipengaruhi beberapa faktor berupa lingkungan, gizi, kependudukan, pendidikan, sosial ekonomi, dan perilaku masyarakat. Diare dapat terjadi pada semua kelompok umur, baik balita, anak-anak, dan orang dewasa dengan berbagai golongan sosial. Diare terkadang dianggap sebagai penyakit ringan, sementara lingkup global dan nasional menunjukkan bahwa diare merupakan satu dari beberapa penyakit mematikan tertinggi di negara miskin dan berkembang, terutama pada balita. Penyakit diare masih menjadi masalah kesehatan masyarakat Indonesia karena morbiditas dan mortalitas yang tinggi (Kementerian Kesehatan RI, 2011).

Infeksi Nosokomial merupakan satu faktor infeksi penyebab diare. Infeksi Nosokomial biasa menyerang bayi atau lansia. Infeksi nosokomial adalah infeksi yang didapat dari rumah sakit ketika pasien sedang dirawat di rumah sakit (Nugraheni & Winarni, 2012). Penelitian ini mengatakan angka kejadian infeksi nosokomial di dunia cukup tinggi, yaitu 5% per tahun atau 9 juta dari 190 juta pasien yang dirawat.

Beberapa faktor risiko infeksi diare nosokomial ialah usia, lama rawat inap, jumlah antibiotik, dan pemberian pipa/selang makanan atau pipa nasogastrik, paparan kemoterapi sitotoksik, penggunaan terapi penekan asam, penghambat

pompa proton (PPIs), kontak dengan sistem perawatan kesehatan (McFarland, 1995).

Penelitian Blot et al. (2003) mengidentifikasi faktor risiko dan karakteristik mikrobiologi pada pasien diare dewasa dan menggunakan faktor seperti konsumsi antibiotik, kemoterapi, terapi radiasi, dan rawat inap selama enam minggu sebelum pemeriksaan tinja lebih lanjut dicatat. Penelitian McFarland (1995) juga mengatakan bahwa penyebab diare nosokomial diketahui bervariasi, seperti bakteri patogen dan paparan selama di rumah sakit, termasuk antibiotik, obat-obatan, dan prosedur medis. Dari 126 kasus diare nosokomial, 37 (29,4%) berasal dari bakteri, 57 (45,2%) berasal dari iatrogenik yang didokumentasikan, dan 32 (25,4%) tidak diketahui asalnya.

Penelitian Kurisu et al. (2019) mengatakan rincian tentang hubungan antara faktor risiko ini dan timbulnya diare nosokomial masih belum jelas. Ada beberapa laporan tentang hubungan antara faktor risiko diare nosokomial dan durasi penggunaan antibiotik, PPI, H2RA, dan pemberian makan tabung. Pengetahuan klinis tentang kepentingan relatif atau pola temporal pengaruh faktor risiko diare nosokomial dapat mengakibatkan meminimalkan pemberian obat-obatan ini atau mempertimbangkan pemberian probiotik untuk mencegah diare nosokomial.

Perkembangan ilmu pengetahuan yang semakin kompleks mendorong manusia untuk mempelajari suatu data yang dapat diolah dan dapat menghasilkan sesuatu. Pada bidang ilmu komputer, terdapat ilmu pembelajaran mesin atau yang biasa disebut machine learning. Machine learning dalam bidang kesehatan dapat digunakan untuk mengklasifikasi penyakit, mengetahui jenis penyakit, serta memprediksi suatu penyakit (Telaumbanua et al., 2020).

Penelitian mengenai faktor risiko diare nosokomial menggunakan algoritma Machine learning sudah pernah dilakukan oleh Kurisu et al., pada tahun 2019. Penelitian tersebut menggunakan data pasien rawat inap dengan infark serebral

akut karena sering memiliki beberapa faktor risiko diare nosokomial, seperti penggunaan PPI, H2RA, dan selang makan.

Metode machine learning yang berpotensi untuk mengklasifikasi suatu penyakit adalah *Support Vector Machine* (SVM). Penelitian ini bertujuan untuk mengetahui hubungan antara faktor risiko dan diare nosokomial dengan menggunakan algoritma Support Vector Machine.

## **1.2 Rumusan Masalah**

Berdasarkan uraian latar belakang tersebut, rumusan masalah penelitian ini adalah klasifikasi diare nosokomial pada lansia menggunakan SVM dan mengevaluasi kinerja SVM.

## **1.3 Tujuan**

Penelitian ini mempunyai beberapa tujuan, yaitu:

1. Mengetahui hubungan antara faktor risiko dan diare nosokomial.
2. Mengetahui kinerja SVM untuk klasifikasi diare nosokomial pada lansia dengan *10 fold cross validation*.

## **1.4 Batasan Masalah**

Penelitian ini mempunyai beberapa batasan masalah, yaitu:

1. Penelitian ini berfokus pada pasien diare lanjut usia.
2. Penelitian ini menggunakan data yang diambil dari link <https://peerj.com/articles/7969/#supplemental-information>.
3. Penelitian ini menggunakan metode SVM dengan 3 *kernel*, yaitu linear, *gaussian*, dan *polynomial*.
4. Penelitian ini menggunakan *10 fold cross validation*.

## **1.5 Manfaat**

Penelitian ini bermanfaat sebagai pengetahuan kinerja SVM dalam menganalisis faktor risiko yang mempengaruhi diare nosokomial. Penelitian ini juga bermanfaat untuk menambah pengetahuan tentang variabel yang berpengaruh penting pada diare nosokomial.

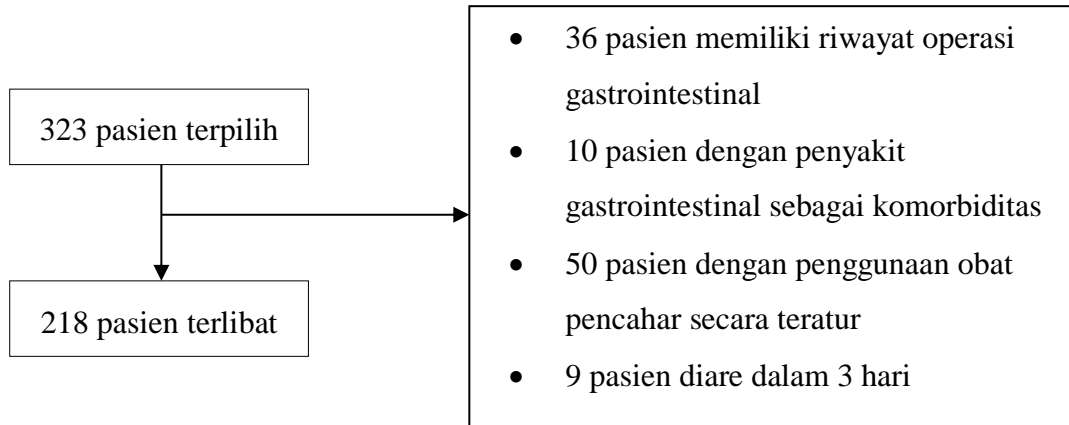
## **II. TINJAUAN PUSTAKA**

### **2.1 Penelitian Terdahulu**

Penelitian terdahulu merupakan penelitian yang sudah ada dengan topik yang sama. Penelitian terdahulu diperlukan guna menambah pengetahuan peneliti. Berikut adalah penelitian terdahulu yang menjadi pendukung penelitian ini.

Kurisu et al. (2019) berhipotesis bahwa algoritma pembelajaran mesin tersebut dapat mengungkapkan hubungan yang tidak diketahui, seperti tingkat kepentingan dan pola temporal pengaruh faktor risiko untuk diare nosokomial, yang akan membantu dokter. Penelitian ini bertujuan untuk mengetahui hubungan antara faktor risiko dan timbulnya diare nosokomial menggunakan algoritma pembelajaran mesin. Penelitian ini menggunakan data pasien rawat inap dengan infark serebral akut karena data tersebut sering memiliki beberapa faktor risiko diare nosokomial, seperti penggunaan PPI, H2RA, dan selang makan.

Penelitian Kurisu et al. (2019) dilakukan di Rumah Sakit Umum Showa, sebuah pusat tersier di Jepang. Penelitian ini mengumpulkan data dari rekam medis elektronik. Penelitian ini mempunyai 323 data yang diambil dari bulan April 2017 sampai Maret 2018. Sebanyak 323 data pasien dengan infark serebral akut berpotensi memenuhi syarat untuk dimasukkan dalam penelitian ini, dan setelah menerapkan kriteria khusus, terdapat 218 pasien dilibatkan. Rincian penerapan kriteria khusus pada penelitian ini dapat dilihat pada Gambar 1.



Gambar 1. Penerapan Kriteria Eksklusi pada Penelitian Terdahulu (Kurusu et al., 2019)

Penelitian Kurisu et al. (2019) menggunakan 218 pasien, 48 mengalami diare selama periode observasi. Pada kelompok diare, tes CDI dilakukan pada 12 (25%) pasien. Di antara mereka, dua (4%) pasien positif untuk antigen GDH dan tidak ada yang positif untuk toksin CD. Waktu timbulnya diare nosokomial berkisar antara 4 sampai 45 hari sejak masuk rumah sakit. Tidak ada data yang hilang untuk setiap variabel. Pasien dalam kelompok diare berusia lanjut, memiliki skor mRS yang lebih tinggi, memiliki tingkat penggunaan antibiotik, selang makan, dan penggunaan H2RA yang lebih tinggi. Mereka juga memiliki waktu penggunaan antibiotik dan selang makan yang lebih lama dibandingkan dengan kelompok non-diare.

Penelitian Kurisu et al. (2019) menggunakan dua metode *machine learning*, yaitu Support Vector Machine dan Random Forest dengan masing-masing menggunakan 5-fold cross validation. Penelitian ini menggunakan kriteria AUC (*Area Under the Curve*) dalam menentukan model yang baik pada klasifikasi studi kasus penelitian ini. Tabel 1 merupakan hasil dari penelitian ini.

Tabel 1. Hasil Penelitian Terdahulu

	AUC	Variabel dihilangkan
<b>5-fold Cross Validation</b>		
Random Forest	0,944	PPI, H2RA
SVM(Gaussian)	0,937	H2RA
SVM(Gaussian)	0,934	Jenis Kelamin
<b>Data Tes</b>		
Random Forest	0,832	PPI, H2RA

Hasil AUC metode *Random Forest* penelitian tersebut adalah 94%. Hasil AUC metode *Support Vector Machine* adalah 93,7% untuk H2RA sebagai variabel yang dihilangkan. Hasil AUC metode *Support Vector Machine* untuk jenis kelamin sebagai variabel yang dihilangkan adalah 93,4%. Hasil AUC data tes pada penelitian tersebut adalah 83,2% untuk metode *Random Forest*.

## 2.2 Diare

Berdasarkan Buletin Jendela Data dan Informasi Kesehatan yang diterbitkan Kementerian Kesehatan RI (2011), Dr. Drg. Magdarina Destri Agniti, MPH, mengatakan diare sebagai penyakit yang disebabkan oleh infeksi mikroorganisme meliputi bakteri, virus, parasit, *protozoa*, dan penularan secara fekal-oral. Pengertian lain dari Prof. dr. Sri Suparyati Soenarto, SpA(K), PhD, mengatakan bahwa diare merupakan suatu kumpulan dari gejala infeksi pada saluran pencernaan yang dapat disebabkan oleh beberapa organisme seperti bakteri, virus, dan parasit. Beberapa organisme tersebut biasanya menginfeksi saluran pencernaan manusia melalui makanan dan minuman yang telah tercemar oleh organisme tersebut (*food borne disease*).

## 2.3 Infeksi Nosokomial

Infeksi nosokomial, juga disebut “Infeksi yang didapat di rumah sakit adalah infeksi yang didapat selama perawatan rumah sakit atau inkubasi saat masuk.. Infeksi yang terjadi lebih dari 48 jam setelah masuk biasanya dianggap nosokomial. Definisi untuk mengidentifikasi infeksi nosokomial telah



dikembangkan untuk situs infeksi tertentu (misalnya saluran kemih, paru). Ini berasal dari yang diterbitkan oleh Pusat Pengendalian dan Pencegahan Penyakit (CDC) di Amerika Serikat atau selama konferensi internasional dan digunakan untuk pengawasan infeksi nosokomial. Mereka didasarkan pada kriteria klinis dan biologis, dan mencakup sekitar 50 lokasi infeksi potensial. Infeksi nosokomial juga dapat dianggap endemik atau epidemi. Infeksi endemik adalah yang paling umum. Infeksi epidemik terjadi selama wabah, yang didefinisikan sebagai peningkatan yang tidak biasa di atas dasar dari infeksi tertentu atau organisme yang menginfeksi (Dellinger, 2016).

Menurut *Center for Disease Control* (2002), pasien dikatakan infeksi nosokomial jika memenuhi beberapa kriteria berikut:

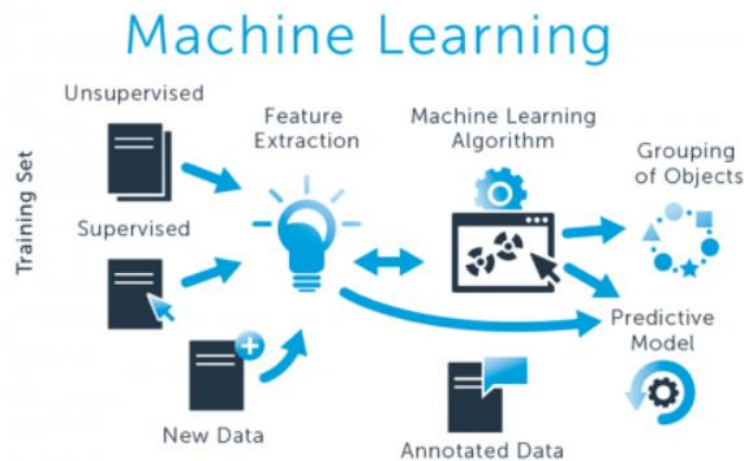
1. Saat penderita mulai dirawat di Rumah Sakit tidak didapatkan tanda-tanda klinis dari infeksi tersebut.
2. Saat penderita mulai dirawat di rumah sakit tidak sedang dalam masa inkubasi dari infeksi tersebut.
3. Tanda-tanda klinis infeksi baru mulai sekurang-kurangnya setelah 3 kali 24 jam sejak mulai perawatan.
4. Infeksi bukan sisa (residual) dari infeksi sebelumnya.

Infeksi nosokomial dapat menyebabkan terjadinya beberapa penyakit seperti gastroenteritis, meningitis, radang paru-paru, dan termasuk diare.

## **2.4 Machine Learning**

*Machine learning* yang berarti pembelajaran mesin telah mengalami kemajuan pesat selama dua dekade terakhir. *Machine learning* adalah mesin yang dikembangkan untuk mempelajari data yang ada dengan menganalisa data tanpa perlu di program kembali oleh penggunanya (Takdirillah, 2020). *Machine Learning* atau pembelajaran mesin merupakan pendekatan *Artificial Intelligence* yang banyak digunakan untuk menggantikan atau menirukan perilaku manusia untuk menyelesaikan masalah atau melakukan otomatisasi (Ahmad, 2017).

Menurut Jordan & Mitchell pada tahun 2015, *Machine learning* merupakan metode pilihan dalam mengembangkan perangkat lunak praktis untuk visi komputer, pengenalan suara, pemrosesan bahasa alami, kontrol robot, dan aplikasi lain. Sebagai bidang studi, *machine learning* merupakan perpaduan antara ilmu komputer, statistik, dan berbagai disiplin ilmu lain yang berkaitan dengan perbaikan otomatis dari waktu ke waktu, serta pengambilan keputusan di bawah ketidakpastian.



Gambar 2. Machine Learning (Dosaer, 2020)

Tujuan machine learning ada dua, yaitu memprediksi masa depan dan memperoleh ilmu pengetahuan (Wira & Putra, 2020). *Machine learning* terbagi menjadi dua jenis, yaitu *supervised learning* dan *unsupervised learning* (Hancock, Zvelebil, & Cristianini, 2004).

### 1. *Supervised Learning*

Metode *supervised learning* merupakan metode yang cepat dan akurat. *Data training* metode ini mencakup *input* dan hasil yang diinginkan. *Supervised learning* didasarkan pada *training* sampel data dari sumber data dengan klasifikasi yang benar yang sudah ditugaskan (Sathya & Abraham, 2013).

### 2. *Unsupervised Learning*

*Unsupervised Learning* digunakan untuk mengelompokkan *input* data dalam kelas berdasarkan sifat statistika saja. *Unsupervised learning* mengacu pada

kemampuan belajar dan mengatur informasi tanpa memberikan sinyal kesalahan untuk mengevaluasi solusi potensial (Sathya & Abraham, 2013).

## 2.5 *Feature Selection*

Selama dekade terakhir, banyak yang menerapkan *feature selection* pada bioinformatika untuk menjadi prasyarat nyata dalam membangun model. Berbeda dengan teknik reduksi lain seperti yang didasarkan pada proyeksi (misalnya analisis komponen utama) atau kompresi (misalnya menggunakan teori informasi), teknik *feature selection* tidak merubah representasi asli dari variabel, tetapi cukup pilih beberapa subset dari mereka. Dengan demikian, mereka mempertahankan semantik asli dari variabel, sehingga dapat menawarkan keuntungan dari interpretabilitas oleh pakar domain (Kursa & Rudnicki, 2010).

*Feature selection* dapat diterapkan pada *supervised* dan *unsupervised learning*. Menurut Putri (2017), tujuan *feature selection* adalah sebagai berikut.

1. Mereduksi dimensi dari fitur yang digunakan pada proses data mining dengan memilih sejumlah fitur yang merupakan subset dari fitur yang lama.
2. Membuat model lebih mudah dipahami serta lebih mudah divisualisasikan.
3. Mengurangi jumlah noise.

*Feature selection* bertujuan mencari parameter model untuk subset fitur yang optimal dan relevan dalam pemodelan tugas. Dalam konteks klasifikasi, teknik *feature selection* dapat diatur ke dalam tiga kategori, tergantung caranya menggabungkan pencarian *feature selection* dengan konstruksi model klasifikasi: metode *filter*, metode pembungkus, dan metode tertanam (Kursa & Rudnicki, 2010).

Beberapa penulis mendefinisikan *feature selection* dari berbagai sudut. Berikut adalah kemampuan *feature selection* yang secara konseptual berbeda dan mencakup berbagai definisi (Dash & Liu, 1997).

1. Menemukan subset fitur berukuran kecil yang diperlukan dan cukup untuk konsep target.
2. Memilih subset fitur sehingga nilai fungsi kriteria dapat dioptimalkan.
3. Memilih subset fitur untuk meningkatkan akurasi prediksi atau mengurangi ukuran struktur tanpa mengurangi secara signifikan akurasi prediksi pengklasifikasi yang dibangun hanya dengan menggunakan fitur yang dipilih.
4. Memilih subset kecil sedemikian rupa sehingga distribusi kelas yang dihasilkan, yang hanya diberi nilai untuk fitur yang dipilih, adalah sedekat mungkin untuk distribusi kelas asli mengingat dari semua nilai fitur.

Idealnya, *feature selection* dilakukan melalui subset fitur, dan mencoba menemukan yang terbaik satu di antara subset kandidat yang bersaing menurut beberapa fungsi evaluasi.

Menurut (Dash & Liu, 1997), ada empat dasar langkah-langkah dalam *feature selection*:

1. Prosedur pembuatan untuk menghasilkan subset kandidat berikutnya;
2. Fungsi evaluasi untuk mengevaluasi subset yang diperiksa;
3. Kriteria berhenti untuk memutuskan kapan harus berhenti; dan
4. Prosedur validasi untuk memeriksa apakah subset tersebut valid

Penelitian ini menggunakan package Boruta dalam pemilihan fitur. Algoritma Boruta adalah package yang dibangun untuk algoritma klasifikasi. Boruta didasarkan pada suatu gagasan, yaitu dengan menambahkan keacakan ke sistem dan mengumpulkan hasil dari kumpulan sampel acak yang dapat mengurangi dampak yang menyesatkan dari fluktuasi acak dan korelasi. Keacakan ekstra ini akan memberi pandangan yang lebih jelas akan atribut yang sangat penting (Kursa & Rudnicki, 2010).

## **2.6 Cross-Validation**

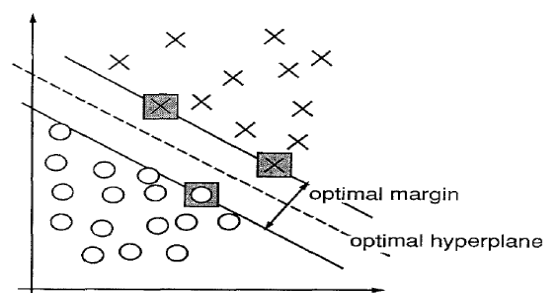
*Cross-validation* merupakan metode statistik untuk mengevaluasi dan membandingkan algoritma pembelajaran. Cara kerja *cross validation* adalah

dengan membagi data menjadi dua segmen, satu digunakan untuk belajar atau melatih model, dan yang lain untuk memvalidasi model. Bentuk dasar *cross-validation* adalah *k-fold cross-validation*. Pada *k-fold cross-validation*, data dipartisi terlebih dahulu menjadi  $k$  segmen atau lipatan berukuran sama. Selanjutnya pelatihan dan validasi  $k$  iterasi dilakukan sedemikian rupa sehingga dalam setiap iterasi lipatan data yang berbeda digunakan untuk validasi (*testing*), sedangkan  $k-1$  lipatan yang tersisa digunakan untuk pembelajaran (*training*) (Refaeilzadeh, Tang, & Liu, 2008).

## 2.7 Support Vector Machine

*Support Vector Machine* (SVM) dikembangkan oleh Boser, Guyon, dan Vapnik tahun 1992 di *Annual Workshop on Computational Learning Theory*. Konsep dasar metode SVM merupakan gabungan atau kombinasi dari teori komputasi yang telah ada sebelumnya, seperti *marginhyperplane*, *kernel* oleh Aronszajn tahun 1950, *Lagrange Multiplier* yang ditemukan oleh Joseph Louis Lagrange tahun 1766, dan demikian dengan konsep pendukung lain (Nugroho, Witarto, & Handoko, 2003).

SVM adalah teknik yang berguna untuk klasifikasi data. Klasifikasi biasanya melibatkan pemisahan data menjadi *data training* dan *data testing*. Setiap contoh dalam *data training* berisi satu nilai target (yaitu label kelas) dan beberapa atribut (yaitu fitur atau variabel yang diamati). Tujuan SVM adalah menghasilkan model (berdasarkan *data training*) yang memprediksi nilai target dari *data testing* (Cortes & Vapnik, 1995).



Gambar 3. Konsep Metode SVM (Cortes & Vapnik, 1995)

Gambar 3 merupakan konsep metode SVM. Algoritma SVM bertujuan untuk menemukan fungsi linear dari bentuk  $f(x) = w^T x + b$ , dengan  $w \in R^p$  dan  $b \in R$  sehingga titik data  $x$  yang telah ditetapkan akan memberikan nilai label +1 jika  $f(x) > 0$ , dan akan memberikan nilai label -1 jika  $f(x) < 0$  (Dong, Xia, & Tu, 2007).

Empat *kernel* dasar yang ada pada SVM (Bhavsar & Panchal, 2012) ditunjukkan pada persamaan (1), (2), (3), dan (4).

1. Linear

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

2. Polynomial

$$K(x_i, x_j) = (\gamma x_i \cdot x_j + r)^d, \gamma > 0 \dots\dots\dots(2)$$

3. Radial Basis Function (RBF), yang biasa dikenal dengan Gaussian

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

4. Sigmoid

$$K(x_i, x_j) = \tanh(K \langle x_i, x_j \rangle + r) \dots\dots\dots(4)$$

Pada hal ini,  $\gamma$ ,  $r$ , dan  $d$  adalah parameter *kernel*. Penelitian ini menggunakan *kernel* linear, *gaussian*, dan *polynomial*.

## 2.8 Confusion Matrix

*Confusion Matrix* berisi informasi tentang klasifikasi aktual dan prediksi yang dilakukan oleh klasifikasi sistem. Performa sistem ini biasanya dievaluasi menggunakan data dalam matriks. Tabel 2 menunjukkan *confusion matrix* untuk klasifikasi dua kelas (Sasongko, 2016).

Tabel 2. *Confusion Matrix*

		Aktual	
		Negatif	Positif
Prediksi	Negatif	TN	FN
	Positif	FP	TP

Dengan penjelasan seperti berikut (Sasongko, 2016):

1. TP adalah model klasifikasi memprediksi dengan benar kelas positif.
2. TN adalah model klasifikasi memprediksi dengan benar kelas negatif.
3. FP adalah model klasifikasi memprediksi dengan salah kelas negatif.
4. FN adalah model klasifikasi memprediksi dengan salah kelas positif.

Pada *confusion matrix*, terdapat beberapa rumus yang dapat digunakan untuk menghitung performa klasifikasi. Penelitian ini menggunakan beberapa kriteria evaluasi seperti *accuracy*, *AUC*, *sensitivity*, dan *specificity*.

1. *Accuracy* adalah presentase jumlah data yang diklasifikasikan dengan benar, baik itu bernilai *true positive* maupun *true negative*. Rumus perhitungan akurasi ditunjukkan pada persamaan 5 (Gorunescu, 2010).

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \times 100\% \quad (5)$$

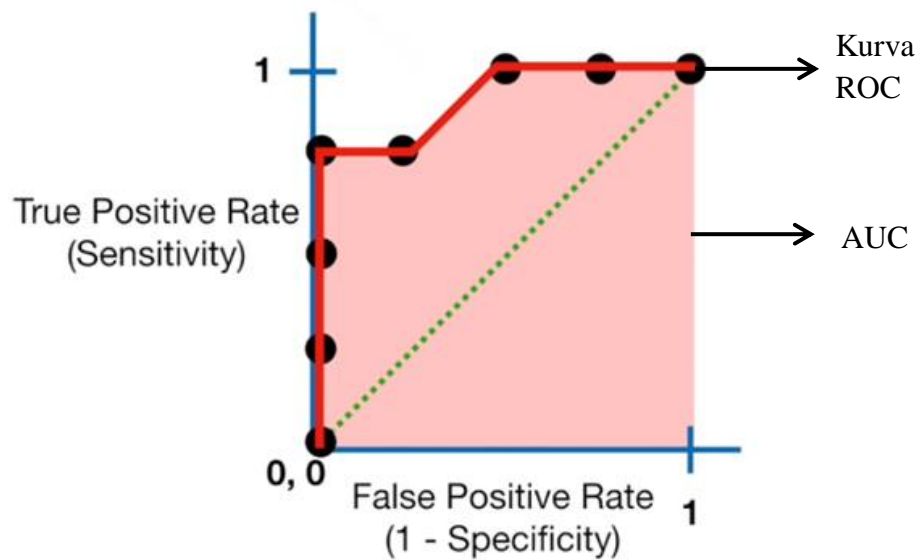
2. *Sensitivity* adalah presentase kasus positif dibandingkan dengan keseluruhan data yang positif dan bernilai *positive* pada hasil prediksi. Rumus *sensitivity* dinyatakan pada persamaan 6 (Gorunescu, 2010).

$$Sensitivity = \frac{TP}{TP+FN} \times 100\% \dots\dots\dots (6)$$

3. *Specificity* adalah presentase kasus negatif dibandingkan dengan keseluruhan data yang negatif dan bernilai *negative* pada hasil prediksi. Rumus *specificity* dinyatakan pada persamaan 7 (Gorunescu, 2010).

$$Specificity = \frac{TN}{TN+FP} \times 100\% \dots\dots\dots (7)$$

4. AUC adalah luas di bawah kurva ROC. AUC merupakan ukuran kinerja yang sering digunakan sebagai ukuran klasifikasi, menggabungkan ambang batas keputusan serta kemiringan kelas dan biaya parameter pada model. AUC dapat membuat model klasifikasi terlihat lebih baik atau lebih buruk dari yang sebenarnya (Flach & Ac, 2011).



Gambar 4. Grafik ROC (Arifin, 2019)

AUC adalah ukuran kinerja yang populer dalam ketidakseimbangan kelas. Semakin tinggi nilai AUC, kinerja yang dihasilkan semakin baik (Press et al., 2013). AUC didapatkan dari rata-rata perkiraan bidang trapesium untuk kurva yang dibuat oleh *True Positive rate* dan *False Positive rate* (Dubey et al., 2014). Rumus AUC dinyatakan pada persamaan 8.

$$\text{AUC} = \frac{1 + \text{TPR} - \text{FPR}}{2} \dots\dots\dots (8)$$

Tabel 3. Kriteria Klasifikasi AUC (Gorunescu, 2010)

Kinerja	Klasifikasi
0,90-1,00	Paling baik
0,80-0,90	Baik
0,70-0,80	Biasa
0,60-0,70	Rendah
0,50-0,60	Gagal

AUC mempunyai kriteria keakuratan klasifikasi seperti pada Tabel 3. Jika kinerja AUC yang dihasilkan dalam rentang 0,90 sampai 1,00 maka kinerja AUC dapat dikatakan paling baik. Jika kinerja AUC yang dihasilkan ada pada rentang 0,80-0,90 maka kinerja AUC dikatakan baik. Selanjutnya jika kinerja AUC ada pada rentang 0,70-0,80 dapat dikatakan biasa, lalu rentang 0,60-0,70 masuk klasifikasi rendah, dan yang terakhir rentang 0,50-0,60 berarti gagal klasifikasi.





No	Kegiatan	Tahun 2020-2021															
		Bulan															
		4	5	6	7	8	9	10	11	12	1	2	3	4			
8	Penulisan bab 4-5																

Berdasarkan Tabel 4, diketahui bahwa proses penulisan bab 1-3 dimulai dari April 2020. Analisis dan pengambilan data dilakukan pada bulan Mei. Setelah data diambil, dilakukan *feature selection*. Setelah itu melakukan *k-Fold Cross Validation* pada bulan September. Proses pemodelan data untuk klasifikasi pada bulan November sampai Desember. Setelah itu, melakukan evaluasi pada Januari 2021 untuk mencari hasil model SVM terbaik.

### 3.2 Data dan Alat

#### 1. Data

Data yang digunakan pada penelitian ini adalah data penderita diare nosokomial pada lansia. Data yang digunakan dalam penelitian ini didapatkan dari jurnal Kurisu et al. pada tahun 2019. Data terdiri dari data pada kelas negatif dan data pada kelas positif. Jumlah data yang terdapat dalam data kelas negatif yaitu berjumlah 170 data dan data kelas positif berjumlah 48 data, sehingga jumlah semua dataset yang digunakan yaitu 218 data. Atribut data terdiri dari umur, jenis kelamin, ranking scale, penghambat pompa proton (PPIs), histamine 2 blocker, antibiotik, selang makan, pembagian kelas diare dan non-diare, dan lama waktu perawatan di rumah sakit. Data pada penelitian ini diperoleh dari alamat website <https://peerj.com/articles/7969/#supplemental-information>

#### 2. Alat

Untuk melakukan penelitian ini, dibutuhkan *hardware* dan beberapa *software* sebagai berikut:

##### a. *Hardware*

Penelitian ini menggunakan *hardware* yaitu laptop dengan spesifikasi *processor* Quad Core A8-7410 APU with AMD Radeon R5 Graphics 2.20

GHz, CPU *up to* 2.5GHz, BIOS X455YI.210, 400 MHz, RAM 4GB, HDD 500GB

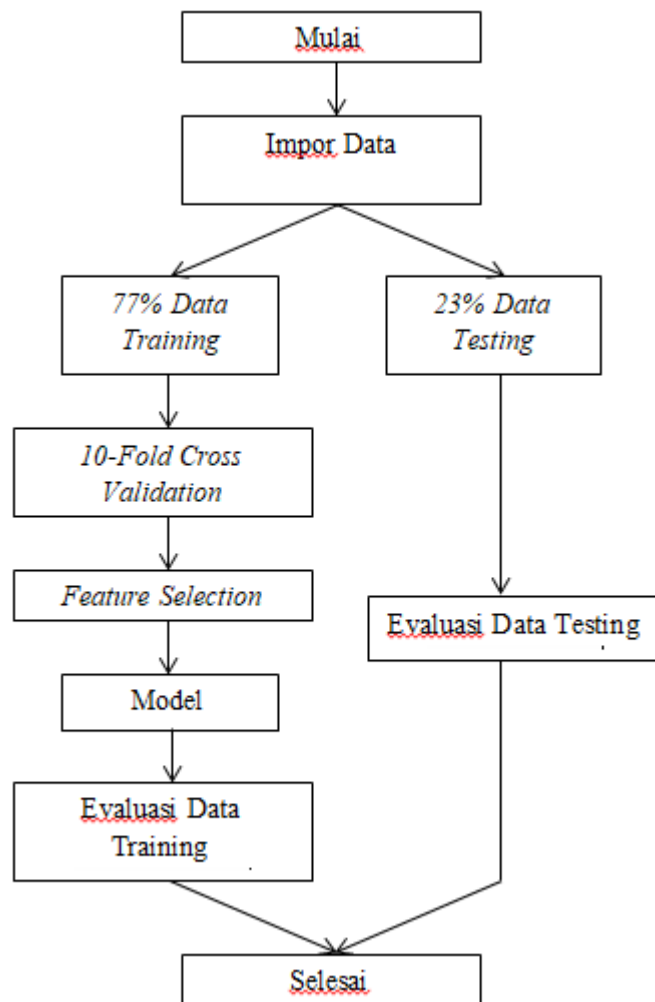
b. *Software*

Beberapa *software* yang digunakan pada penelitian ini, yakni sistem operasi Windows 10 Profesional 64-bit, Google Chrome untuk mengakses dan mengambil data, Microsoft Excel 2010 untuk menampilkan data, R studio for Windows versi 3.5.1 sebagai bahasa pemrograman, dan beberapa *package* yang digunakan pada penelitian ini, seperti:

- 1) *Package readr* versi 1.3.1 (James Hester, 21 Desember 2018)
- 2) *Package caret* versi 6.0-86 (Max Kuhn, 20 Maret 2020)
- 3) *Package e1071* versi 1.7-3 (David Meyer, 27 November 2019)
- 4) *Package Boruta* versi 7.0.0 (Miron Bartosz Kurska, 2010)
- 5) *Package PROC* versi 1.16.2 (Xavier Robin, 20 Maret 2020)

### **3.3 Alur Kerja Penelitian**

Alur kerja penelitian adalah langkah yang dilakukan peneliti dalam melakukan penelitian. Alur kerja penelitian ini terbagi menjadi beberapa tahapan, seperti yang ada pada Gambar 5.

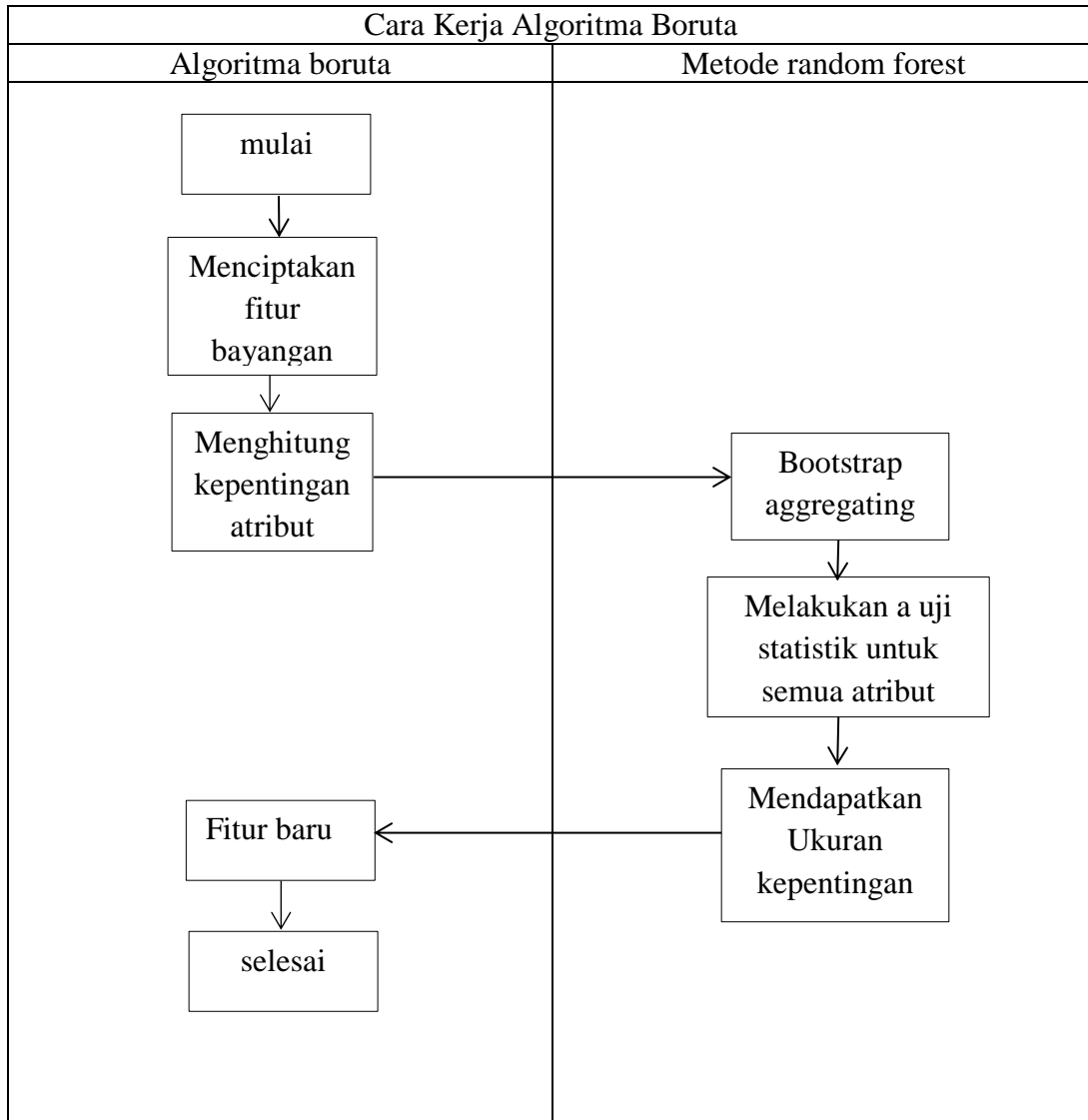


Gambar 5. Alur Kerja Penelitian

Penjelasan metode implementasi yang dilakukan pada penelitian ini adalah sebagai berikut:

#### 1. *Feature Selection*

*Feature selection* adalah teknik pada *machine learning* untuk mendapatkan model yang lebih baik dengan cara memilih fitur yang penting dari semua fitur yang ada. Penelitian ini menggunakan package boruta pada pelaksanaan feature selection. Gambar 6 menunjukkan cara kerja algoritma package boruta.



Gambar 6. Cara Kerja Algoritma Boruta

Algoritma boruta dimulai dengan menciptakan fitur bayangan dengan cara melakukan pengacakan dari fitur yang sudah ada sehingga terbentuk frame baru disebut  $X_{Boruta}$  yang dapat dilihat pada Gambar 7.

	age	height	weight		income
<b>0</b>	25	182	75	<b>0</b>	20
<b>1</b>	32	176	71	<b>1</b>	32
<b>2</b>	47	174	78	<b>2</b>	45
<b>3</b>	51	168	72	<b>3</b>	55
<b>4</b>	62	181	86	<b>4</b>	61

X and y

Gambar 7. Ilustrasi Contoh Dataset

Selanjutnya  $X_{\text{Boruta}}$  dimodelkan dengan  $y(\text{income})$  menggunakan random forest untuk menghitung kepentingan masing-masing atribut. Random forest akan melakukan bootstrap aggregating, yaitu teknik yang banyak diterapkan pada algoritma klasifikasi yang bertujuan meningkatkan akurasi pengklasifikasi dengan menggabungkan pengklasifikasi tunggal dan mempunyai hasil lebih baik daripada random sampling. Subset dari data awal akan dimodelkan dengan decision tree hingga mendapatkan satu hasil. Dilakukan hal yang sama untuk perulangan kedua hingga membentuk banyak decision tree. Perulangan tersebut akan memberikan informasi mengenai fitur yang mempunyai kepentingan paling tinggi hingga paling rendah. Hal itu disebut sebagai ambang batas yang digunakan untuk melihat tingkat kepentingan suatu fitur. Setelah itu, maksimum ambang batas dari fitur bayangan dijadikan threshold. Selanjutnya dilakukan uji statistik untuk semua atribut atau yang biasa disebut dengan iterasi. Setelah melakukan iterasi, akan didapatkan ukuran kepentingan dan ukuran tersebut digunakan untuk menentukan fitur baru yang akan digunakan pada percobaan dengan feature selection.

## 2. *k-Fold Cross Validation*

Tahap ini mulai dilakukan pembagian data training dan data testing. Penelitian ini menggunakan  $k=10$  untuk pembagian data. Masing-masing pembagian berupa  $(k-1)/k$  untuk data training dan  $1/k$  untuk data testing.

### 3. Pemodelan dengan SVM

Tahap ini mulai dilakukan pembagian *training* SVM pada *data training* dengan proses pembelajaran untuk model SVM. Hal yang terlebih dahulu dilakukan yaitu dengan menentukan tipe *kernel* dan nilai parameter. Jika model data sudah terpilih maka sudah dapat melakukan proses klasifikasi.

### 4. Evaluasi

Tahap terakhir adalah tahap evaluasi. Setelah data sudah diklasifikasikan menggunakan SVM dan sudah dilakukan 10-Fold Cross Validation, maka hasil dari penelitian tersebut akan dilakukan evaluasi untuk menguji hasil yang sudah dilakukan. Kriteria evaluasi yang digunakan yaitu menggunakan *confusion matrix* yang terdiri dari *Accuracy*, *AUC*, *Sensitivity*, dan *Specificity*. Selanjutnya yaitu, menghitung rata-rata *Accuracy* dan *AUC* yang sudah dilakukan sebanyak 10- *Fold Cross Validation* untuk mendapatkan hasil kinerja klasifikasi diare nosokomial pada lansia menggunakan Support Vector Machine.

## V. SIMPULAN DAN SARAN

### 5.1 Simpulan

Simpulan dari penelitian yang dilakukan adalah sebagai berikut:

1. Faktor risiko penting yang berhubungan dengan diare nosokomial adalah antibiotik, umur, jenis kelamin, *ranking scale*, selang makan, dan lama waktu perawatan di rumah sakit. Sedangkan faktor risiko yang kurang penting pada penelitian ini adalah variabel PPIs dan *h2 blocker*.
2. Kinerja metode *Support Vector Machine* (SVM) dengan *feature selection* dan 10-fold cross validation untuk klasifikasi diare nosokomial pada lansia dapat membuat model semakin lebih baik, dilihat dari hasil nilai rata-rata AUC pada penelitian ini.
  - a. Kernel gaussian pada metode SVM memiliki hasil kinerja terbaik dengan AUC tertinggi dibandingkan dengan kernel Polynomial dan Linear pada metode SVM. Pada penelitian ini metode SVM (*Support Vector Machine*) kernel gaussian memiliki nilai AUC sebesar 95,30%.
  - b. Dengan menggunakan 10-fold cross validation, kernel Gaussian menjadi metode yang tepat untuk klasifikasi diare nosokomial.

### 5.2 Saran

Adapun saran yang diberikan untuk penelitian selanjutnya mengenai Klasifikasi Diare Nosokomial Pada Lansia Menggunakan *Support Vector Machine* ini yaitu sebagai berikut :

1. Pada penelitian selanjutnya diharapkan agar dapat mencari *feature selection* yang lebih bagus guna meningkatkan akurasi.



2. Pada penelitian selanjutnya diharapkan dapat menggunakan metode yang tidak menggunakan feature selection seperti Deep Learning sehingga dapat mengetahui kinerja dan hasil akurasi dengan metode yang berbeda.

## DAFTAR PUSTAKA

- Ahmad, A. 2017. *Mengenal Artificial Intelligence, Machine Learning, Neural Network, dan Deep Learning*. (October).
- Arifin, Rio. 2019. <https://arifinrio95.medium.com/memahami-roc-dan-auc-2e0e4f3638bf>. Diakses pada 18 Mei 2021.
- 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.
- Blot, E., Escande, M. C., Besson, D., Barbut, F., Granpeix, C., Asselain, B., ... Pouillart, P. 2003. Outbreak of Clostridium difficile-related diarrhoea in an adult oncology unit: Risk factors and microbiological characteristics. *Journal of Hospital Infection*, 53(3), 187–192. <https://doi.org/10.1053/jhin.2002.1356>
- Center for Disease Control. 2002. Overview of CDC's campaign to prevent antimicrobial resistance in health settings. <https://www.cdc.gov/media/pressrel/r020326.htm>. Diakses pada 28 Maret 2021.
- Cortes, C., & Vapnik, V. 1995. Dispensing system boosts throughput 50 percent. *Assembly*, 44(13), 97.
- Dash, M., & Liu, H. 1997. Feature selection for classification. *Intelligent Data Analysis*, 1(3), 131–156. <https://doi.org/10.3233/IDA-1997-1302>
- Dellinger, E. P. 2016. Prevention of Hospital-Acquired Infections. *Surgical Infections*, 17(4), 422–426. <https://doi.org/10.1089/sur.2016.048>
- Dong, Y., Xia, Z., & Tu, M. 2007. Selecting Optimal Parameters in Support Vector Machines. *Proceedings of the IEEE 6th International Conference on Machine Learning and Applications (ICMLA07)*, (July).
- Dosaer, Erik. 2020. <http://www.postmedya.com/default/mengenal-lebih-dalam-tentang-machine-learning/>. Diakses pada 18 Mei 2021
- Dubey, R., Zhou, J., Wang, Y., Thompson, P. M., & Ye, J. 2014. Analysis of sampling techniques for imbalanced data: An n=648 ADNI study.

- NeuroImage, 87, 220–241. <https://doi.org/10.1016/j.neuroimage.2013.10.005>
- Flach, P., Flach, P., & Ac, B. 2011. A Coherent Interpretation of AUC as a Measure of Aggregated Classification Performance.
- Gorunescu, F. 2010. *Intelligent Systems Reference Library Volume 12*.
- Hancock, J. M., Zvelebil, M. J., & Cristianini, N. 2004. Supervised and Unsupervised Learning. *Dictionary of Bioinformatics and Computational Biology*, (April). <https://doi.org/10.1002/9780471650126.dob0716.pub2>
- Jordan, M. I., & Mitchell, T. M. 2015. *Science-ML-2015*. 349(6245).
- Kementerian Kesehatan RI. 2019. *Profil Kesehatan Indonesia 2018 [Indonesia Health Profile 2018]*. Retrieved from [http://www.depkes.go.id/resources/download/pusdatin/profil-kesehatan-indonesia/Data-dan-Informasi\\_Profil-Kesehatan-Indonesia-2018.pdf](http://www.depkes.go.id/resources/download/pusdatin/profil-kesehatan-indonesia/Data-dan-Informasi_Profil-Kesehatan-Indonesia-2018.pdf)
- Kementerian Kesehatan RI. 2011. Buletin Jendela Data dan Informasi Kesehatan Situasi Diare di Indonesia. *Jurnal Buletin Jendela Data & Informasi Kesehatan*, 2, 1–44.
- Kurusu, K., Yoshiuchi, K., Ogino, K., & Oda, T. 2019. Machine learning analysis to identify the association between risk factors and onset of nosocomial diarrhea: A retrospective cohort study. *PeerJ*, 2019(10), 1–15. <https://doi.org/10.7717/peerj.7969>
- Kursa, M. B., & Rudnicki, W. R. 2010. Feature Selection with the Boruta Package. *Journal of Statistical Software*, 36(11), 1–13. <https://doi.org/10.18637/jss.v036.i11>
- McFarland, L. V. 1995. Epidemiology of infectious and iatrogenic nosocomial diarrhea in a cohort of general medicine patients. *AJIC: American Journal of Infection Control*, 23(5), 295–305. [https://doi.org/10.1016/0196-6553\(95\)90060-8](https://doi.org/10.1016/0196-6553(95)90060-8)
- Nugraheni, R., & Winarni, S. 2012. Infeksi Nosokomial di RSUD Setjonegoro Kabupaten Wonosobo. *Media Kesehatan Masyarakat Indonesia*, 11(1), 94–100. <https://doi.org/10.14710/mkmi.11.1.94-100>
- Nugroho, A. S., Witarto, A. B., & Handoko, D. 2003. Application of Support Vector Machine in Bioinformatics. *Proceeding of Indonesian Scientific Meeting in Central Japan*.
- Press, I., Shafer, L., Arnold, G. W., & Jacobson, D. 2013. Imbalanced Learning. *Imbalanced Learning*. <https://doi.org/10.1002/9781118646106>
- Putri, L. A. A. R. (2017). Seleksi Fitur Dalam Klasifikasi Genre Musik. *Jurnal*

*Ilmiah Ilmu Komputer Udayana*, 10(1), 19–26.

- Refaeilzadeh, P., Tang, L., & Liu, H. 2008. Cross-Validation. Cross-Validation.
- 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. *International Journal of Advanced Research in Artificial Intelligence*, 2(2), 34–38. <https://doi.org/10.14569/ijarai.2013.020206>
- Takdirillah, Robby. 2020. Apa itu Machine Learning? Beserta Pengertian dan Cara Kerjanya. <https://www.dicoding.com/blog/machine-learning-adalah/>. Diakses pada 18 Mei 2021.
- Telaumbanua, F. D., Hulu, P., Nadeak, T. Z., Lumbantong, R. R., & Dharma, A. 2020. Penggunaan Machine Learning Di Bidang Kesehatan. *Jurnal Teknologi Dan Ilmu Komputer Prima (JUTIKOMP)*, 2(2), 57–64. <https://doi.org/10.34012/jutikomp.v2i2.657>
- Wira, J., & Putra, G. 2020. *Pengenalan Konsep Pembelajaran Mesin dan Deep Learning*. 4.