PENERAPAN *ADAPTIVE BOOSTING-SUPPORT VECTOR MACHINE* PADA KLASIFIKASI PASIEN KANKER PARU-PARU BERDASARKAN FAKTOR RISIKO

SKRIPSI

OLEH
AULIA NURSAFITRI
NIM. 220601110087

PROGRAM STUDI MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM MALANG
2026

PENERAPAN *ADAPTIVE BOOSTING-SUPPORT VECTOR MACHINE* PADA KLASIFIKASI PASIEN KANKER PARU-PARU BERDASARKAN FAKTOR RISIKO

SKRIPSI

Diajukan Kepada
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang
untuk Memenuhi Salah Satu Persyaratan dalam
Memperoleh Gelar Sarjana Matematika (S.Mat)

Oleh
Aulia Nursafitri
NIM. 220601110087

PROGRAM STUDI MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM MALANG
2026

PENERAPAN *ADAPTIVE BOOSTING-SUPPORT VECTOR MACHINE* PADA KLASIFIKASI PASIEN KANKER PARU-PARU BERDASARKAN FAKTOR RISIKO

SKRIPSI

Oleh
Aulia Nursafitri
NIM. 220601110087

Telah Disetujui Untuk Diuji

Malang, 28 April 2026

Dosen Pembimbing I	Dosen Pembimbing II
Ria Dhea Layla Nur Karisma, M.Si	Juhari, M.Si
NIPPPK. 19900709 202321 2 037	NIPPPK. 19840209 202321 1 010

Mengetahui,
Ketua Program Studi Matematika

Dr. Fachrur Rozi, M.Si
NIP. 19800527 200801 1 012

PENERAPAN *ADAPTIVE BOOSTING-SUPPORT VECTOR MACHINE* PADA KLASIFIKASI PASIEN KANKER PARU-PARU BERDASARKAN FAKTOR RISIKO

SKRIPSI

Oleh
Aulia Nursafitri
NIM. 220601110087

Telah Dipertahankan di Depan Dewan Penguji Skripsi
dan Dinyatakan Diterima sebagai Salah Satu Persyaratan
untuk Memperoleh Gelar Sarjana Matematika (S.Mat)

Tanggal 11 Mei 2026

Ketua Penguji	:	Dr. Fachrur Rozi, M.Si.	.....
Anggota Penguji 1	:	Abdul Aziz, M.Si.	.....
Anggota Penguji 2	:	Ria Dhea Layla Nur Karisma, M.Si.	.....
Anggota Penguji 3	:	Juhari, M.Si.	.....

Mengetahui,
Ketua Program Studi Matematika

KEMENTERIAN AGAMA ★ FAKULTAS SAINS DAN TEKNOLOGI ★ UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM MALANG ★ PROGRAM STUDI MATEMATIKA ★ REPUBLIK INDONESIA

Dr. Fachrur Rozi, M.Si
NIP. 19800527 200801 1 012

PERNYATAAN KEASLIAN TULISAN

Saya bertanda tangan di bawah ini

Nama	:	Aulia Nursafitri
NIM	:	220601110087
Program Studi	:	Matematika
Fakultas	:	Sains dan Teknologi
Judul Skripsi	:	Penerapan *Adaptive Boosting-Support Vector Machine* pada Klasifikasi Pasien Kanker Paru-Paru Berdasarkan Faktor Risiko.

Menyatakan dengan sebenarnya bahwa skripsi yang saya tulis ini merupakan hasil karya sendiri, bukan pengambilan tulisan atau pemikiran orang lain yang saya akui sebagai pemikiran saya, kecuali dengan mencantumkan sumber cuplikan pada daftar rujukan di halaman terakhir. Apabila dikemudian hari terbukti skripsi ini adalah hasil jiplakan atau tiruan, maka saya bersedia menerima sanksi yang berlaku atas perbuatan tersebut.

Malang, 18 Mei 2026

Aulia Nursafitri
NIM. 220601110087

MOTO

"Orang yang bersungguh-sungguh untuk mencari keridaan Allah, niscaya Allah akan menunjukkan jalan baginya"
(QS. Al-'Ankabut 29:69)

"*Everything you lose is a step you take*."
(Taylor Swift)

HALAMAN PERSEMBAHAN

Dari sekian banyak lembar skripsi, lembar persembahan menjadi bagian paling bermakna bagi penulis. Dengan penuh rasa syukur dan terima kasih, karya sederhana ini penulis persembahkan kepada:

1. Pelindung sekaligus harta paling berharga penulis, kedua orang tua tercinta, Bapak Agus Irawan dan Ibu Sri Mulyani yang tidak pernah berhenti memberikan kasih sayang dan melangitkan doa-doa baik untuk penulis. Adik penulis tersayang, Fakhri, yang selalu menghadirkan canda tawa dan warna di hidup penulis. Tiada kata yang cukup untuk membalas segala cinta dan perjuangan yang telah diberikan. Semoga karya sederhana ini menjadi salah satu bentuk kebanggaan untuk segala pengorbanan yang begitu besar.
2. Salwa, Dian, Kailla, Rahma, Lala, Eti, dan Shafira sahabat kandung penulis. Terima kasih karena selalu menjadi tempat pulang dan sumber motivasi bagi penulis. Perjuangan dan semangat kalian yang tidak pernah berhenti dalam menyelesaikan skripsi menjadi dorongan besar bagi penulis untuk terus bertahan dan menyelesaikan perjalanan ini.
3. Sahabat dan teman seperjuangan penulis, terima kasih telah menjadi bagian dari perjalanan perkuliahan penulis. Kehadiran kalian selalu menjadi pengalaman terbaik yang akan penulis ingat, meskipun jarak akan memisahkan kita nanti. Semoga selalu ada kesempatan untuk kembali bertemu, berbagi cerita, dan merayakan kesuksesan bersama di kemudian hari.
4. Terakhir, untuk diri saya sendiri. Terima kasih telah bertahan, berjuang, dan tidak menyerah meskipun banyak rasa cemas, takut, lelah, dan keraguan yang datang silih berganti. Terima kasih untuk setiap langkah keberanianmu dalam menyelesaikan perjalanan ini. Semoga segala proses yang dilalui menjadi awal dari langkah-langkah baik di pengalaman selanjutnya.

KATA PENGANTAR

Segala puji bagi Allah SWT. yang telah memberikan limpahan rahmat dan karunia-Nya sehingga penulis mampu menyelesaikan skripsi yang berjudul "Penerapan *Adaptive Boosting-Support Vector Machine* pada Klasifikasi Pasien Kanker Paru-Paru Berdasarkan Faktor Risiko". Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana Matematika di Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.

Penulis menyadari bahwa selama proses penyusunan skripsi ini, penulis banyak mendapatkan bantuan, bimbingan, dan dukungan dari banyak pihak. Oleh karena itu, dengan hormat penulis ingin menyampaikan rasa terima kasih yang sebesar-besarnya kepada:

1. Prof. Dr. Hj. Ilfi Nur Diana, M.Si., CAHRM., CRMP, selaku Rektor Universitas Islam Negeri Maulana Malik Ibrahim Malang.
2. Dr. H. Agus Mulyono, S,Pd., M.Kes., selaku Dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.
3. Dr. Fachrur Rozi, M.Si., selaku Ketua Program Studi Matematika, Universitas Islam Negeri Maulana Malik Ibrahim Malang sekaligus Ketua Penguji dalam Ujian Skripsi penulis yang telah memberikan kritik dan saran dalam penulisan skripsi penulis.
4. Ria Dhea Layla Nur Karisma, M.Si., selaku Dosen Pembimbing I yang selalu memberikan bimbingan, arahan, serta dukungan dalam penyelesaian skripsi penulis.
5. Juhari, M.Si., selaku Dosen Pembimbing II yang telah memberikan arahan, nasihat, dukungan, dan motivasi dalam penyelesaian skripsi penulis.
6. Abdul Aziz, M.Si., selaku Anggota Penguji 1 dalam Ujian Skripsi yang telah memberikan kritik dan saran dalam penulisan skripsi penulis.
7. Seluruh Dosen Program Studi Matematika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.
8. Ayah dan Ibu tercinta yang tiada hentinya memberikan kasih sayang, doa, dukungan, motivasi, dan keyakinan bahwa penulis pasti bisa menyelesaikan

apapun yang telah penulis mulai, serta Adik tersayang yang selalu menghibur penulis dalam penyusunan skripsi penulis.

9. Karina Ayu dan sahabat-sahabat seperjuangan penulis yang senantiasa selalu menyemangati dan menghibur penulis selama proses penyusunan skripsi penulis.
10. Teman-teman PKL Taspen yang selalu membersamai suka duka dalam pengerjaan skripsi penulis.
11. Seluruh teman-teman Program Studi Matematika Angkatan 2022 yang selalu memberikan motivasi dan semangat kepada penulis dalam proses menyelesaikan skripsi penulis.

Semoga Allah SWT. membalas segala bantuan dan kebaikan yang telah diberikan kepada penulis. Penulis sadar bahwa penyusunan skripsi ini masih memiliki kekurangan. Penulis berharap skripsi ini dapat bermanfaat bagi penulis dan pembaca serta memberikan wawasan untuk referensi penelitian selanjutnya.

Malang, 18 Mei 2026

Penulis

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

DAFTAR SIMBOL

X_1	:	Usia Pasien
X_2	:	Jenis Kelamin Pasien
X_3	:	Status Merokok
X_4	:	Komorbiditas
X_5	:	Riwayat Keluarga
Y_1	:	Stadium Kanker Paru-Paru
$\boldsymbol{x}_i$	:	Vektor data ke-i
y_i	:	Label kelas dari data ke-i
$(\boldsymbol{x}_n, y_n)$	:	Pasangan data latih dengan n adalah jumlah sampel
$\boldsymbol{w}$	:	Vektor bobot
b	:	Bias
ρ	:	Jarak terdekat antara batas pemisah dengan vektor data tiap kelas
L	:	Fungsi lagrangian
α_i	:	*Lagrange Multiplier*
ξ_i	:	*Slack variable*
C	:	Parameter regulasi untuk keseimbangan *margin*
$\varphi(\boldsymbol{x})$	:	Fungsi pemetaan
$\mathbb{R}$	:	Bilangan *real*
$K(\boldsymbol{x}_i, \boldsymbol{x}_j)$	:	Fungsi kernel antara data ke-i dan ke-j
c	:	Konstanta
d	:	Derajat pada kernel polinomial
γ	:	Parameter kernel RBF
σ	:	Parameter kernel *sigmoid*
$D_1(i)$	:	bobot data ke-i pada iterasi pertama
ε_t	:	Tingkat kesalahan
α_t	:	Bobot model klasifikasi lemah pada iterasi ke-t
$h_t(\boldsymbol{x})$	:	Model SVM pada iterasi ke-t
$D_{t+1}(i)$	:	Bobot baru pada iterasi selanjutnya
Z_t	:	Faktor normalisasi bobot pada iterasi ke-t
T	:	Jumlah iterasi maksimum pada algoritma AdaBoost
$G(\boldsymbol{x})$	:	Hipotesis kuat atau model akhir klasifikasi
E_i	:	*Error* pada lipatan ke-i
F_i	:	Lipatan ke-i
Z_i	:	Data terstandardisasi
$h(\boldsymbol{x})$	:	Model SVM
χ^2	:	*Chi-square*
$f(\boldsymbol{w})$	:	Fungsi objektif
T'	:	Iterasi yang memenuhi $\varepsilon_t > 0.5$

DAFTAR LAMPIRAN

ABSTRAK

Nursafitri, Aulia. 2026. **Penerapan *Adaptive Boosting-Support Vector Machine* pada Klasifikasi Pasien Kanker Paru-Paru Berdasarkan Faktor Risiko**. Skripsi. Program Studi Matematika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang.
Pembimbing: (1) Ria Dhea Layla Nur Karisma, M.Si. (2) Juhari, M.Si.

Kata kunci: Klasifikasi, *Support Vector Machine*, *Adaptive Boosting*, Kanker Paru-Paru, Faktor Risiko

Penelitian ini menerapkan algoritma gabungan *Support Vector Machine* (SVM) dengan *Adaptive Boosting* (AdaBoost) untuk mengklasifikasikan pasien kanker paru-paru berdasarkan faktor risiko dengan SVM sebagai *base learner*. SVM bekerja dengan mencari *hyperplane* optimal untuk memisahkan data ke dalam dua kelas, yaitu kelas positif dan kelas negatif. Sementara itu, AdaBoost digunakan untuk meningkatkan performa model dengan menggabungkan beberapa model sederhana menjadi model yang lebih akurat melalui pembaruan bobot pada setiap iterasi. Data yang digunakan merupakan data sekunder pasien kanker paru-paru tahun 2020-2025 yang diperoleh dari rekam medis RSUD Karsa Husada Batu. Tahap analisis meliputi penanganan *missing value* menggunakan imputasi modus, standardisasi data, pembagian data *training* dan data *testing*, serta pelatihan model menggunakan kernel linear dengan parameter terbaik $cost(C) = 10$ dan melakukan boosting sebanyak 10 iterasi. Evaluasi model dilakukan menggunakan 5-*Fold Cross Validation* dan *Confusion Matrix*. Hasil evaluasi 5-*Fold Cross Validation* memperoleh nilai rata-rata akurasi sebesar 89.96% yang menunjukkan tingkat kestabilan model. Selain itu, hasil evaluasi *Confusion Matrix* memperoleh nilai akurasi sebesar 76.92% yang menunjukkan bahwa model AdaBoost-SVM mampu mengklasifikasikan sebagian besar pasien kanker paru-paru dengan benar. Selanjutnya, nilai *recall* sebesar 50%, presisi sebesar 33.33%, dan F1-*Score* sebesar 40%. Nilai metrik evaluasi tersebut dipengaruhi oleh kondisi data yang tidak seimbang (*data imbalanced*) di mana jumlah pasien stadium IV lebih banyak dibandingkan pasien stadium III.

ABSTRACT

Nursafitri, Aulia. 2026. **Application of Adaptive Boosting-Support Vector Machine in the Classification of Lung Cancer Patients Based on Risk Factors.** Undergraduate Thesis. Mathematics Study Program, Faculty of Science and Technology, Universitas Islam Negeri Maulana Malik Ibrahim Malang.
Advisors: (1) Ria Dhea Layla Nur Karisma, M.Si. (2) Juhari, M.Si.

Keywords: Classification, Support Vector Machine, Adaptive Boosting, Lung Cancer, Risk Factor

This study applied a combined algorithm of Support Vector Machine (SVM) with Adaptive Boosting (AdaBoost) to classify lung cancer patients based on risk factors, with SVM as the base learner. SVM works by finding the optimal hyperplane to separate the data into two classes: positive and negative. Meanwhile, AdaBoost improves model performance by combining multiple simple models into more accurate ones through weight updates at each iteration. The data used is secondary data on lung cancer patients in 2020-2025, obtained from the medical records of Karsa Husada Batu Hospital. The analysis stage includes handling missing values with mode imputation, data standardization, splitting the data into training and test sets, and training a linear kernel model with $C = 10$ and boosting for up to 10 iterations. Model evaluation was performed using 5-fold cross-validation and a Confusion Matrix. The results of the 5-Fold Cross Validation evaluation yielded an average accuracy of 89.96%, indicating the model's stability. In addition, the Confusion Matrix evaluation yielded an accuracy of 76.92%, indicating that the AdaBoost-SVM model can classify most lung cancer patients. Furthermore, the recall value is 50%, the precision is 33.33%, and the F1-Score is 40%. The evaluation metric was influenced by an imbalance in the data, with more stage IV patients than stage III patients.

مستخلص البحث

نور سفطري ، أوليا ٢٠٢٦. **تطبيق تعزيز التكيف – آلات المتجهات الداعمة في تصنيف مرضى سرطان الرئة بناءً على عوامل الخطر** .البحث الجامعي .قسم الرياضيات، كلية العلوم والتكنولوجيا، جامعة مولانا مالك إبراهيم الإسلامية الحكومية مالانج. المشرف: (١) ريا ديّا ليلى نور كاريسما، الماجستير في العلوم. (٢) جوهري، الماجستير في العلوم.

الكلمات الأساسية :التصنيف، آلات المتجهات الداعمة، تعزيز التكيف، سرطان الرئة، عوامل الخطر.

طبقت هذه الدراسة خوارزمية هجينة تجمع بين آلة المتجهات الداعمة (*SVM*) وتقنية التعزيز التكيفي (*AdaBoost*)لتصنيف مرضى سرطان الرئة بناءً على عوامل الخطر، حيث استُخدمت خوارزمية *SVM* بوصفها المتعلم الأساسي .(*base learner*) تعمل خوارزمية *SVM* على إيجاد المستوى الفاصل الأمثل (*hyperplane*) لفصل البيانات إلى فئتين، هما الفئة الإيجابية والفئة السلبية، في حين تُستخدم *AdaBoost* لتحسين أداء النموذج من خلال تحديث الأوزان في كل تكرار. اعتمدت الدراسة على بيانات ثانوية لمرضى سرطان الرئة خلال الفترة *2020–2025*، والتي تم الحصول عليها من السجلات الطبية لمستشفى .*RSUD Karsa Husada Batu* شملت مراحل التحليل معالجة القيم المفقودة باستخدام تعويض المنوال، وتوحيد البيانات، وتقسيم البيانات إلى بيانات تدريب واختبار، بالإضافة إلى تدريب النموذج باستخدام النواة الخطية (*linear kernel*) مع أفضل قيمة للمعلمة $C = 10$ وتنفيذ عملية التعزيز لمدة 10 تكرارات. تم تقييم النموذج باستخدام أسلوب التحقق المتقاطع بخمس طيات(5-*Fold Cross Validation*) ومصفوفة الالتباس .(*Confusion Matrix*) أظهرت نتائج التحقق المتقاطع متوسط دقة بلغ 89.96% مما يدل على استقرار أداء النموذج، كما أظهرت نتائج مصفوفة الالتباس دقة مقدارها 76.92% مما يشير إلى قدرة نموذج *AdaBoost-SVM* على تصنيف معظم مرضى سرطان الرئة. إضافة إلى ذلك، بلغت قيمة الاسترجاع50% (*Recall*) ، والدقة النوعية 33.33% (*Precision*)، ودرجة F1-*Score* مقدارها 40%. وقد تأثرت قيم مقاييس التقييم بحالة عدم توازن البيانات(*imbalanced data*) ، حيث كان عدد مرضى المرحلة الرابعة أكبر من عدد مرضى المرحلة الثالثة.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Support Vector Machine (SVM) adalah metode *machine learning* berjenis *supervised learning* di mana algoritma biasanya dilatih menggunakan data berlabel untuk menangani masalah klasifikasi dan prediksi. Algoritma ini dapat diterapkan dalam berbagai konteks, yaitu kategorisasi gambar, klasifikasi teks, mengidentifikasi spam, pengenalan wajah, dan mendeteksi anomali (Abushark dkk., 2025). *Support Vector Machine* (SVM) diperkenalkan oleh Vladimir Vapnik pada tahun 1995. Menurut Cortes & Vapnik (1995), SVM memiliki konsep melakukan klasifikasi dua kelas data, yaitu kelas positif dan kelas negatif dengan mencari garis pemisah (*hyperplane*) yang optimal. *Hyperplane* yang optimal dapat diperoleh dari *margin* atau jarak antara *hyperplane* dengan data terdekat di masing-masing kelas. SVM menjadi algoritma efektif karena memiliki ketahanan terhadap *overfitting*. Meskipun demikian, SVM masih memiliki kekurangan, seperti kendala dalam menangani data kompleks, pemilihan parameter yang kurang tepat, dan ketidakseimbangan data yang berpengaruh pada akurasi model (Maulana dkk., 2025). Kekurangan tersebut dapat diatasi dengan mengoptimalkan performa SVM menggunakan algoritma *ensemble learning*.

Ensemble learning merupakan metode *machine learning* yang memiliki konsep menggabungkan beberapa model yang berbeda dengan tujuan memperoleh hasil yang lebih akurat dibanding menggunakan satu model saja. *Ensemble learning* memiliki konsep utama yaitu *boosting* yang bekerja dengan mengoptimalkan model

sederhana untuk meningkatkan akurasi model. *Ensemble learning* dapat menjadi fondasi dari metode *boosting* seperti *Adaptive Boosting* (Schapire, 1990).

Adaptive Boosting (AdaBoost) adalah salah satu metode *machine learning* yang bekerja dengan melatih model sederhana secara berulang pada nilai bobot yang diperbarui di setiap iterasi. Konsep AdaBoost yaitu memberikan bobot awal yang sama pada seluruh data latih. Setelah model sederhana diperoleh, bobot akan diperbarui sesuai dengan hasil klasifikasi pada tiap iterasi. Beberapa model sederhana tersebut akan digabungkan menjadi satu model yang akurat. AdaBoost menjadi salah satu metode *ensemble* yang paling efektif karena berusaha mengurangi kesalahan dengan memfokuskan pembelajaran pada data yang salah atau sulit untuk diklasifikasikan (Freund & Schapire, 1995).

Penggabungan AdaBoost dengan SVM memanfaatkan SVM sebagai *base learner* yang menghasilkan model klasifikasi di setiap iterasi. AdaBoost bekerja dengan mengatur proses pelatihan SVM secara berulang melalui pembaruan bobot berdasarkan hasil klasifikasi. Tujuan dari pembaruan bobot tersebut yaitu supaya algoritma lebih fokus pada data yang salah diklasifikasikan pada iterasi sebelumnya di mana data yang salah klasifikasi akan diberikan bobot yang lebih besar. Proses ini dilakukan hingga mencapai iterasi maksimum (T). Hasil dari beberapa model sederhana dari SVM akan digabungkan dan membentuk satu model yang kuat (Wang, 2012).

Penelitian pada penerapan AdaBoost dan SVM telah banyak dilakukan dalam berbagai kasus. Penelitian oleh Syafika & Karisma (2023) menerapkan SVM untuk mengklasifikasikan indeks khusus penanganan stunting. Kernel yang digunakan yaitu kernel polinomial dengan parameter $h = 1$ dan $C = 100$. Hasil klasifikasi

memperoleh tingkat akurasi sebesar 100% menunjukkan SVM mampu mengklasifikasikan dengan sangat baik.

Penelitian oleh Sinaga dkk. (2022) menerapkan gabungan AdaBoost-*Random Forest* untuk mendeteksi kanker paru-paru. *Random Forest* berperan sebagai *base learner*. Hasil penelitian menunjukkan algoritma AdaBoost-*Random Forest* mampu menghasilkan akurasi yang lebih baik yaitu 95.4% dibandingkan hanya menggunakan *Random Forest* tunggal.

Penelitian oleh Listiana & Muslim (2017) menerapkan AdaBoost-SVM untuk mengklasifikasikan diagnosa *chronic kidney disease*. Penelitian ini menggunakan kernel linear dengan parameter C yaitu 10. Penelitian ini menunjukkan AdaBoost mampu meningkatkan akurasi SVM sebesar 99.5% dibanding hanya SVM tunggal sebesar 62.5%.

Penelitian oleh Adyalam dkk. (2018) menerapkan AdaBoost-SVM untuk mengklasifikasikan tingkat keparahan penyakit *osteoarthritis*. Penelitian ini menerapkan SVM menggunakan kernel RBF dan data *training* 10% hingga 90%. Hasil penelitian ini menunjukkan bahwa AdaBoost dapat meningkatkan nilai akurasi model SVM mencapai 85.714% dengan data *training* 80% dibanding hanya menggunakan SVM tunggal

Penelitian oleh Nabeel dkk. (2024) menerapkan beberapa algoritma yang dioptimalkan dengan *hyperparameter tuning* untuk mengklasifikasikan kanker paru-paru. Pada algoritma SVM digunakan kernel RBF. Hasil penelitian menunjukkan bahwa SVM memperoleh akurasi mencapai 99.16% setelah dioptimalkan dengan *hyperparameter tuning*.

Banyak penderita kanker paru-paru mengalami telat diagnosis di mana sel kanker sudah menyebar ke organ tubuh lainnya (Alfarisa dkk., 2023). Menurut Putri dkk. (2024), biasanya kanker paru-paru mudah menyerang seseorang berusia 55 tahun tanpa gejala awal yang dirasakan. Namun, semakin lama akan menimbulkan gejala umum, seperti batuk parah yang berkepanjangan, sesak nafas, sering merasakan nyeri pada dada, batuk berdarah, pneumonia, batuk disertai pilek, dan gejala lainnya. Gejala fisik yang terjadi yaitu menurunnya berat badan tanpa sebab yang jelas dan mudah merasa lelah. Kanker paru-paru seringkali baru terdeteksi ketika stadium lanjut, yaitu stadium III dan IV di mana sel kanker sudah menyebar ke organ lain yang menyebabkan pasien hanya memiliki peluang hidup sekitar 1-5 persen untuk bertahan hidup. Sedangkan, jika sel kanker dapat terdeteksi lebih awal di stadium I dan II, pasien akan memiliki peluang hidup lebih besar sekitar 40-50 persen untuk bertahan hidup hingga lima tahun (Rejeki & Pratiwi, 2020).

Banyak cara untuk mencegah terdiagnosis kanker paru-paru, salah satunya dengan menghindari pola hidup yang tidak sehat. Sebagaimana Allah SWT. berfirman dalam QS. Al-Baqarah ayat 195 (Kemenag, 2024):

"*..., janganlah jerumuskan dirimu ke dalam kebinasaan, dan berbuat baiklah. Sesungguhnya Allah menyukai orang-orang yang berbuat baik.*" (Q.S. Al-Baqarah: 195).

Berdasarkan ayat di atas, Allah SWT. memerintahkan manusia untuk tidak menjerumuskan diri ke dalam kebinasaan. Manusia dilarang untuk melakukan perbuatan atau kebiasaan yang dapat merugikan diri sendiri dan dianjurkan untuk selalu berbuat baik (Shihab, 2002). Hal tersebut menyadarkan kita untuk tidak mendekati faktor-faktor yang menjerumuskan kita ke dalam suatu penyakit, salah satunya kanker paru-paru.

Berdasarkan data *World Health Organization* (WHO), pada tahun 2022 terdapat hampir 2.5 juta jiwa terdiagnosis kanker paru-paru dan lebih dari 1.8 juta jiwa diantaranya tidak dapat bertahan hidup akibat penyakit tersebut. Kanker paru-paru menjadi penyakit dengan angka kematian terbesar yaitu mencapai 13% diantara penyakit kanker lainnya (Almaidah & Ambarwati, 2022). Tingginya angka kematian ini menyadarkan kita seberapa pentingnya upaya untuk mendeteksi kanker paru-paru lebih awal dengan menghindari faktor-faktor yang berisiko. Oleh karena itu, penelitian klasifikasi kanker paru-paru berdasarkan faktor risikonya akan berguna sebagai informasi penting bagi tenaga medis dalam menentukan strategi penanganan dan upaya menyadarkan masyarakat untuk menghindari faktor yang memicu kanker paru-paru.

Penelitian ini dilakukan dengan memanfaatkan data laporan riwayat kesehatan atau rekam medis setiap pasien kanker paru-paru di RSUD Karsa Husada Batu. RSUD tersebut memiliki sejarah sebagai Rumah Sakit Paru Batu pada tahun 2007 yang kemudian berganti nama menjadi RSUD Karsa Husada Batu pada tahun 2015. Oleh karena itu, penelitian ini memilih RSUD Karsa Husada Batu dengan harapan dapat memperoleh data pasien kanker paru-paru yang memadai.

Penelitian ini bertujuan untuk menerapkan algoritma *Adaptive Boosting-Support Vector Machine* dalam mengklasifikasikan pasien kanker paru-paru berdasarkan faktor risiko yang dialami pasien tersebut. Penelitian ini diharapkan mampu membangun model klasifikasi yang akurat dan bermanfaat bagi tenaga medis dan masyarakat mengenai deteksi awal kanker paru-paru berdasarkan faktor risikonya.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, dapat diambil perumusan masalah dalam penelitian ini sebagai berikut:

1. Bagaimana hasil klasifikasi menggunakan model AdaBoost-SVM dalam mengklasifikasikan pasien kanker paru-paru?
2. Bagaimana tingkat performa dari model klasifikasi AdaBoost-SVM dalam mengklasifikasikan pasien kanker paru-paru?

1.3 Tujuan

Berdasarkan rumusan masalah, tujuan dari penelitian ini sebagai berikut:

1. Mengetahui hasil klasifikasi menggunakan model AdaBoost-SVM dalam mengklasifikasikan pasien kanker paru-paru.
2. Mengukur tingkat performa dari model klasifikasi AdaBoost-SVM dalam mengklasifikasikan kanker paru-paru.

1.4 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

1. Bagi Program Studi Matematika dan Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang
   a. Memberikan bahan literatur dan referensi akademik untuk dikembangkan oleh mahasiswa atau peneliti lain terutama dalam bidang statistika.
   b. Membantu menjalin kemitraan dengan RSUD Karsa Husada Batu untuk memberikan peluang kepada mahasiswa atau peneliti lain untuk melakukan penelitian di instansi terkait.

2. Bagi RSUD Karsa Husada
    a. Membantu rumah sakit dalam memanfaatkan data rekam medis untuk mengklasifikasikan kanker paru-paru berdasarkan faktor risiko untuk mendukung upaya deteksi dini agar masyarakat menghindari faktor-faktor yang menyebabkan terdiagnosis kanker paru-paru.
    b. Memberikan kontribusi dalam membangun reputasi instansi sebagai rumah sakit yang mendukung penelitian mahasiswa.
3. Bagi Masyarakat
    a. Memberikan informasi kepada masyarakat mengenai faktor risiko kanker paru-paru sehingga dapat lebih peduli terhadap upaya deteksi dini kanker paru-paru.
    b. Meningkatkan kesadaran kepada masyarakat untuk meningkatkan kualitas hidup menjadi lebih sehat sebagai upaya mencegah terdiagnosis kanker paru-paru.

1.5 Batasan Masalah

Batasan masalah dalam penelitian ini sebagai berikut:

1. Jenis kernel dibatasi menggunakan kernel linear karena menyesuaikan ukuran dataset yang terbatas.
2. Penelitian menggunakan data rekam medis pasien kanker paru-paru dari RSUD Karsa Husada Batu tahun 2020-2025.
3. Penelitian hanya menggunakan data pasien kanker paru-paru stadium III dan IV karena menyesuaikan ketersediaan data di RSUD Karsa Husada Batu.
4. Performa kinerja dibatasi pada hasil evaluasi model menggunakan K-*Fold Cross Validation* dan *Confusion Matrix*.

5. Penelitian ini difokuskan pada model yang tidak hanya memiliki nilai akurasi yang tinggi, tetapi juga mampu mendeteksi kelas negatif dan kelas positif.

1.6 Definisi Istilah

Beberapa istilah yang digunakan dalam penelitian ini sebagai berikut:

Machine Learning	:	Bagian dari kecerdasan buatan komputer yang mampu bekerja tanpa mendapatkan perintah manusia.
Data *Mining*	:	Teknik analisis untuk menemukan informasi melalui ekstraksi data dari sekumpulan basis data yang besar.
Hyperplane	:	Garis atau batas pemisah yang membedakan dua kelas data dalam SVM, yaitu kelas positif dan negatif.
Margin	:	Jarak terdekat antara batas pemisah dengan data dari masing-masing kelas.
Quadratic Programming	:	Metode untuk mengoptimalkan nilai panjang dari vektor bobot dengan kendala linear.
Lagrange Multiplier	:	Metode untuk menentukan nilai minimum dan maksimum suatu fungsi.
Slack Variable	:	Variabel tambahan untuk memberikan toleransi pada model untuk menangani

		data yang sulit atau mengandung kesalahan.
Kernel	:	Fungsi matematika yang membantu memetakan data ke dimensi yang lebih tinggi.
Dot product	:	Teknik mengalikan dua vektor atau lebih.
Jarak *Euclidean*	:	Jarak antara dua titik atau dua garis lurus yang digunakan pada rumus kernel RBF.
Weak learner	:	Model klasifikasi yang memiliki performa hanya sedikit lebih baik dari prediksi acak.
Strong learner	:	Model klasifikasi yang memiliki performa lebih tinggi dan memberikan nilai prediksi yang sangat akurat.
Iterasi	:	Proses pengujian yang dilakukan secara berulang-ulang sampai mendapatkan hasil yang optimum.
Weighted error	:	Ukuran tingkat kesalahan pada model klasifikasi.
Histologi kanker	:	Jenis sel yang tampak di bawah mikroskop.
Komorbiditas	:	Penyakit yang dimiliki selain diagnosis utama.

BAB II

KAJIAN TEORI

2.1 Teori Pendukung

2.1.1 *Preprocessing* Data

Preprocessing data merupakan tahapan penting dalam proses data *mining* untuk menyiapkan data supaya lebih siap untuk digunakan dengan optimal oleh algoritma. Tahapan ini bertujuan untuk mengubah data mentah yang tidak teratur menjadi lebih teratur dan siap diproses oleh algoritma. Apabila data tidak disiapkan akan mengakibatkan algoritma data *mining* mengalami kesalahan saat diproses maupun memberikan hasil yang tidak akurat (García dkk., 2015). Menurut García dkk. (2015), tahapan *preprocessing* data sebagai berikut:

1. Pelabelan Data: Data yang digunakan akan diberikan label atau kategori tertentu pada variabel yang ingin diklasifikasikan.
2. Pembersihan Data: Data yang digunakan akan dilakukan pembersihan data dari data duplikasi maupun *missing value*. Penanganan *missing value* dilakukan menggunakan teknik imputasi data yang memiliki beberapa mekanisme *missing value* sebagai berikut (Widyawati dkk., 2025):
   a. *Missing Completely At Random* (MCAR)

      Kondisi ketika data yang hilang terjadi secara benar-benar acak dan tidak dipengaruhi oleh variabel lain maupun variabel itu sendiri. Metode sederhana seperti *mean imputation* dan sejenisnya masih dapat digunakan dan menghasilkan model yang valid.

b. *Missing At Random* (MAR)

Kondisi ketika data yang hilang memiliki pola atau berkaitan dengan variabel lainnya. Karena adanya pola antar variabel, metode penanganan yang disarankan yaitu imputasi ganda dan imputasi multivariat.

c. *Missing Not At Random* (MNAR)

Kondisi ketika data yang hilang memiliki pola dengan variabel itu sendiri sehingga sulit diamati dan ditangani secara statistik. Pada kondisi ini, mekanisme tidak dapat diidentifikasi secara pasti dan keberadaannya hanya dapat diduga secara logis karena *missing value* tidak tersedia dalam dataset.

Pengujian mekanisme *missing value* dilakukan dengan membandingkan kelompok data yang memiliki *missing value* dan data yang tidak memiliki *missing value* menggunakan variabel indikator, yaitu bernilai 1 jika data *missing* dan bernilai 0 jika data tidak *missing*. Selanjutnya, dilakukan pengujian untuk mengetahui hubungan antara indikator *missing* dan variabel penelitian. Jika tidak terdapat hubungan, maka *missing value* dapat diindikasikan terjadi secara acak sepenuhnya (MCAR) (Nicholson dkk., 2016). Pengujian tersebut dapat menggunakan uji *Chi-Square* atau *Fisher's Exact* untuk variabel kategorik dan uji *Mann-Whitney* untuk variabel numerik.

Uji *Chi-Square* digunakan untuk menguji hubungan antara indikator *missing value* dan variabel kategorik dengan membandingkan frekuensi pengamatan dan frekuensi harapan. Pengujian *Chi-Square* harus memiliki minimal 80% sel dengan frekuensi harapan lebih dari 5 dan tidak ada sel

dengan frekuensi harapan kurang dari 1. Jika syarat tersebut tidak terpenuhi, maka dapat digunakan uji *Fisher's Exact* (Mchugh, 2013). Hipotesis yang digunakan yaitu:

H_o: Tidak terdapat hubungan antara variabel penelitian yang diuji.

H_1: Terdapat hubungan antara variabel penelitian yang diuji.

Statistik uji *Chi-Square* dapat dinyatakan pada Persamaan (2.1).

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} \tag{2.1}$$

dengan O_i merupakan frekuensi pengamatan pada sel ke-i dan E_i adalah frekuensi harapan pada sel ke-i. Derajat kebebasan dapat dihitung menggunakan rumus:

$$df = (r-1)(c-1) \tag{2.2}$$

dengan r adalah jumlah baris dan c adalah jumlah kolom pada tabel kontingensi. Kriteria keputusan dilakukan dengan membandingkan nilai statistik uji χ^2_{hitung} dengan χ^2_{tabel} atau berdasarkan nilai $p - value$. Jika $\chi^2_{hitung} \leq \chi^2_{tabel}$ atau nilai $p - value > 0.05$, maka terima H_0. Sedangkan, jika $\chi^2_{hitung} > \chi^2_{tabel}$ atau $p - value \leq 0.05$, maka tolak H_0.

Uji *Fisher's Exact* merupakan alternatif pengujian apabila syarat uji *Chi-Square* tidak terpenuhi, yaitu ketika terdapat lebih dari 20% sel memiliki frekuensi harapan kurang dari 5. Uji ini digunakan untuk menguji hubungan antara dua variabel kategorik pada tabel kontingensi 2×2. Nilai $p - value$ pada uji *Fisher's Exact* diperoleh dari perhitungan statistik peluang eksak dari tabel yang diamati untuk menentukan ada atau tidaknya hubungan antar

variabel menggunakan rumus pada Persamaan (2.3) (Sukmana & Rozi, 2017).

$$P = \frac{(A+B)!\,(C+D)!\,(A+C)!\,(B+D)!}{N!\,A!\,B!\,C!\,D!} \tag{2.3}$$

dengan $A, B, C,$ dan D adalah frekuensi pada masing-masing sel tabel kontingensi, sedangkan N adalah jumlah seluruh pengamatan. Kriteria pengambilan keputusan dilakukan berdasarkan nilai $p - value$. Jika nilai $p - value > 0.05$, maka terima H_0, sedangkan jika nilai $p - value \leq 0.05$, maka tolak H_0.

Uji *Mann-Whitney* digunakan untuk membandingkan dua kelompok dengan tujuan mengetahui apakah terdapat perbedaan distribusi nilai variabel antar kelompok tersebut. Uji ini dapat digunakan pada data berskala numerik ketika asumsi normalitas tidak terpenuhi (Wahyudin dkk., 2022). Hipotesis yang digunakan yaitu:

H_o: Tidak terdapat perbedaan distribusi nilai variabel antara kedua kelompok.

H_1: Terdapat perbedaan distribusi nilai variabel antara kedua kelompok.

Uji *Mann Whitney* dilakukan dengan membandingkan distribusi nilai variabel antar dua kelompok yang dapat diurutkan berdasarkan peringkat atau *ranking*. Jika kedua kelompok memiliki distribusi nilai yang sama, maka setiap nilai pada kedua kelompok memiliki peluang yang sama untuk lebih besar atau lebih kecil sehingga tidak terdapat perbedaan distribusi pada kedua kelompok. Statistik uji *Mann-Whitney* dinyatakan sebagai berikut:

$$U_x = n_x n_y + \left(\frac{n_x(n_x+1)}{2}\right) - R_x \tag{2.4}$$

$$U_y = n_x n_y + \left(\frac{n_y(n_y+1)}{2}\right) - R_y \quad (2.5)$$

dengan n_x merupakan jumlah pengamatan pada kelompok pertama, n_y merupakan jumlah pengamatan pada kelompok kedua, R_x merupakan jumlah peringkat pada kelompok pertama, dan R_y merupakan jumlah peringkat pada kelompok kedua. Setelah diperoleh U_x dan U_y, maka nilai statistik uji *Mann-Whitney* yang digunakan adalah nilai terkecil dari kedua uji statistik tersebut. Kriteria pengambilan keputusan dapat dilakukan menggunakan nilai statistik U untuk sampel kecil atau berdasarkan nilai $p - value$. Jika $U_{hitung} \geq U_{tabel}$ atau $p - value > 0.05$, maka terima H_o. Jika $U_{hitung} < U_{tabel}$ atau $p - value < 0.05$, maka tolak H_o. Apabila ukuran sampel pada masing-masing kelompok lebih dari 20, maka distribusi statistik uji akan mendekati distribusi normal sehingga pengujian dapat dilakukan menggunakan nilai statistik Z pada Persamaan (2.6).

$$Z = \frac{U - \mu_U}{\sigma_U} \quad (2.6)$$

dengan $\mu_U = \frac{n_x n_y}{2}$ dan $\sigma_U = \sqrt{\frac{n_x n_y (N+1)}{12}}$, di mana $N = (n_x + n_y)$. Dengan demikian, pengambilan keputusan dapat dilakukan dengan membandingkan nilai statistik uji $|Z_{hitung}|$ dengan nilai Z_{tabel} atau berdasarkan nilai $p - value$. Jika $|Z_{hitung}| \leq Z_{\alpha/2}$ atau $P - value > 0.05$, maka terima H_o. Sedangkan jika Jika $|Z_{hitung}| > Z_{\alpha/2}$ atau $P - value \leq 0.05$, maka tolak H_o.

3. Standardisasi Data: Teknik mengubah nilai numerik dalam data supaya memiliki skala yang sama. Pada penelitian ini, standardisasi data dilakukan

menggunakan *Z-Score* di mana setiap nilai data akan diubah berdasarkan rata-rata dan standar deviasi. Tujuannya adalah supaya data memiliki rata-rata adalah nol dan standar deviasi adalah satu. Rumus *Z-Score* dapat dinyatakan pada Persamaan (2.7) (Ranganathan dkk., 2018):

$$Z_i = \frac{\mathrm{X}_i - \overline{\mathrm{X}}}{s} \tag{2.7}$$

di mana:

Z_i : Nilai terstandardisasi ke-i,

X_i : Nilai data ke-i,

$\overline{\mathrm{X}}$: Nilai rata-rata,

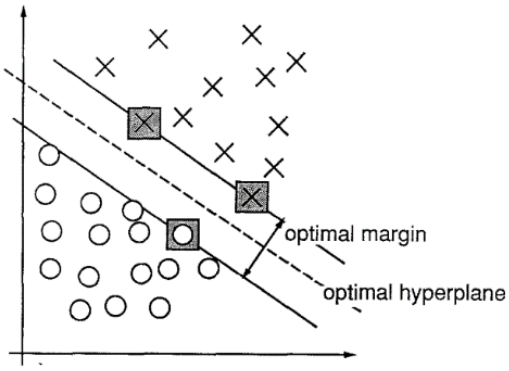
s : Standar deviasi.

Setelah *preprocessing* selesai dilakukan, data akan siap digunakan untuk analisis data *mining*. Data yang dipilih akan digunakan untuk menemukan informasi penting sesuai kebutuhan menggunakan algoritma tertentu.

2.1.2 *Support Vector Machine* (SVM)

Support Vector Machine (SVM) adalah algoritma *supervised learning* yang umumnya digunakan untuk menangani masalah klasifikasi dan regresi baik linear maupun nonlinear. SVM mengklasifikasikan data ke dalam dua kelas, yaitu kelas positif (1) dan kelas negatif (-1). Proses klasifikasi dilakukan dengan mencari batas pemisah (*hyperplane*) terbaik, yaitu *margin* (jarak) yang paling dekat dengan data pada tiap kelas. Pada tahap pelatihan, digunakan data berlabel untuk menemukan *hyperplane* terbaik, kemudian model yang dihasilkan akan diuji pada data baru yang tidak berlabel. Hasil prediksi klasifikasi model akan dibandingkan

dengan label sebenarnya untuk menilai keakuratan model (Mahendro & Abimanto, 2023).



Sumber:(Cortes & Vapnik, 1995)
Gambar 2.1 Ilustrasi SVM

SVM terdiri dari SVM linear dan SVM nonlinear. SVM linear adalah metode klasifikasi yang digunakan ketika data dapat dipisahkan secara mudah oleh *hyperplane*. Menurut Cortes & Vapnik (1995), SVM menggunakan data latih $(\boldsymbol{x_1}, y_1), \ldots, (\boldsymbol{x_n}, y_n)$ dengan n adalah jumlah sampel yang terdiri dari vektor data $(\boldsymbol{x_i})$ dan label kelas (y_i). Setiap vektor data $(\boldsymbol{x_i})$ memiliki label kelas (y_i). SVM bertujuan mencari *hyperplane* terbaik untuk mengklasifikasikan dua kelas data. Persamaan *hyperplane* dapat dituliskan sebagai berikut (Cortes & Vapnik, 1995):

$$\boldsymbol{w} \cdot \boldsymbol{x} + b = 0, \tag{2.8}$$

dengan $\boldsymbol{w}$ adalah vektor bobot dan b adalah bias. Persamaan tersebut akan terpenuhi dengan syarat masing-masing kelas, yaitu:

$$\boldsymbol{w} \cdot \boldsymbol{x_i} + b > 1 \quad \text{jika} \quad y_i = 1 \tag{2.9}$$

$$\boldsymbol{w} \cdot \boldsymbol{x_i} + b < -1 \quad \text{jika} \quad y_i = -1 \tag{2.10}$$

Margin merupakan jarak terdekat antara *hyperplane* dan vektor data dari tiap kelas. *Margin* dapat dirumuskan dengan (Cortes & Vapnik, 1995):

$$\rho(\boldsymbol{w}, b) = \min_{\{\boldsymbol{w}:y=1\}} \frac{\boldsymbol{x} \cdot \boldsymbol{w}}{|\boldsymbol{w}|} - \max_{\{\boldsymbol{w}:y=-1\}} \frac{\boldsymbol{x} \cdot \boldsymbol{w}}{|\boldsymbol{w}|} \tag{2.11}$$

di mana $|\boldsymbol{w}|$ adalah panjang dari vektor bobot $\boldsymbol{w}$. *Hyperplane* yang optimal diperoleh dengan memaksimalkan jarak (*margin*) *hyperplane* antar kelas yang dapat dituliskan dengan Persamaan (2.12).

$$\rho(\boldsymbol{w}, \boldsymbol{b}) = \frac{2}{|\boldsymbol{w}|} \tag{2.12}$$

Margin maksimum dapat diperoleh melalui proses optimasi menggunakan *Quadratic Programming* (QP) dengan meminimalkan nilai $|\boldsymbol{w}|$. Dengan demikian, fungsi objektif untuk memaksimalkan margin dapat ditulis sebagai berikut:

$$f(\boldsymbol{w}) = \frac{1}{2}|\boldsymbol{w}|^2 \tag{2.13}$$

dengan syarat

$$y_i(\boldsymbol{w} \cdot \boldsymbol{x_i} + b) \geq 1, \forall_i \tag{2.14}$$

di mana $y_i \in \{-1,1\}$.

Margin maksimum pada persamaan (2.13) dapat diselesaikan menggunakan *Lagrange Multiplier*. *Lagrange Multiplier* dapat dirumuskan sebagai:

$$L = \frac{1}{2}|\boldsymbol{w}|^2 - \sum_{i=1}^{n} \alpha_i [y_i(\boldsymbol{w} \cdot \boldsymbol{x_i} + b) - 1] \tag{2.15}$$

dengan α_i adalah *Lagrange Multiplier* untuk data ke-i yang berhubungan dengan $\boldsymbol{x}_i$. Nilai α_i harus bernilai lebih dari atau sama dengan nol di mana jika $\alpha_i = 0$ maka data tidak mempengaruhi *hyperplane* dan jika $\alpha_i > 0$ maka data merupakan *support vector*. Kemudian, *L* diminimalkan terhadap $\boldsymbol{w}$ dan *b* sehingga diperoleh:

$$\frac{\partial L}{\partial \boldsymbol{w}} = \mathbf{0}$$

$$\boldsymbol{w} - \sum_{i=1}^{n} \alpha_i y_i \boldsymbol{x_i} = \mathbf{0}$$

$$\boldsymbol{w} = \sum_{i=1}^{n} \alpha_i y_i \boldsymbol{x_i} \tag{2.16}$$

$$\frac{\partial L}{\partial b} = 0$$

$$\sum_{i=1}^{n} \alpha_i y_i = 0 \tag{2.17}$$

Proses memaksimalkan L terhadap α_i melalui substitusi persamaan (2.16) dan (2.17) ke persamaan (2.15) sehingga diperoleh persamaan (2.18) sebagai berikut:

$$L = \frac{1}{2}|\boldsymbol{w}|^2 - \sum_{i=1}^{n} \alpha_i [y_i(\boldsymbol{w} \cdot \boldsymbol{x_i} + b) - 1]$$

$$= \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j (\boldsymbol{x_i} \cdot \boldsymbol{x_j}) - \sum_{i=1}^{n} \alpha_i [y_i(\boldsymbol{w} \cdot \boldsymbol{x_i} + b) - 1]$$

$$= \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j (\boldsymbol{x_i} \cdot \boldsymbol{x_j}) - \sum_{i=1}^{n} \alpha_i \left[y_i \left(\sum_{j=1}^{n} \alpha_j y_j (\boldsymbol{x_j} \cdot \boldsymbol{x_i}) + b \right) - 1 \right]$$

$$= \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j (\boldsymbol{x_i} \cdot \boldsymbol{x_j}) - \sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j (\boldsymbol{x_i} \cdot \boldsymbol{x_j}) + b \sum_{i=1}^{n} \alpha_i y_i + \sum_{i=1}^{n} \alpha_i$$

$$= \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j (\boldsymbol{x_i} \cdot \boldsymbol{x_j}) \tag{2.18}$$

di mana $\alpha_i \geq 0$, $\sum_{i=1}^{n} \alpha_i y_i = 0$.

Umumnya, kelas pada SVM sulit terpisahkan dengan sempurna oleh *hyperplane* sehingga syarat (2.14) tidak terpenuhi. Oleh karena itu, *soft margin*

dipergunakan dengan melibatkan *slack variable* yaitu $\xi_i > 0$ pada persamaan (2.14), sehingga diperoleh:

$$y_i(\boldsymbol{w} \cdot \boldsymbol{x_i} + b) \geq 1 - \xi_i \tag{2.19}$$

Dengan demikian, persamaan (2.13) dapat diformulasikan kembali menjadi,

$$f(\boldsymbol{w}) = \frac{1}{2}|\boldsymbol{w}|^2 + C\sum_{i=1}^{n}\xi_i \tag{2.20}$$

dengan ξ_i digunakan untuk meminimalkan kesalahan dalam klasifikasi. Parameter C digunakan untuk mengatur keseimbangan antara *margin* maksimum dengan kesalahan klasifikasi. Nilai C yang besar akan menghasilkan model yang kompleks dan toleransi kesalahan yang rendah, sedangkan nilai C yang kecil akan menghasilkan model sederhana dan toleransi kesalahan yang tinggi.

Berdasarkan hasil optimasi *margin*, diperoleh model SVM linear yang digunakan untuk proses klasifikasi. Model tesebut dapat dituliskan sebagai berikut (Cortes & Vapnik, 1995):

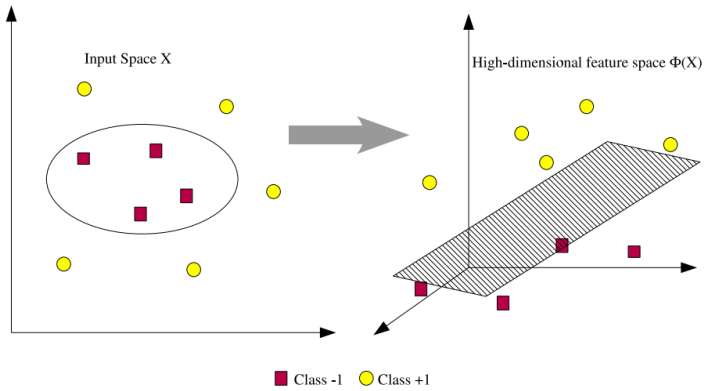
$$\begin{aligned} h(\boldsymbol{x}) &= sign(\boldsymbol{w} \cdot \boldsymbol{x} + b) \\ &= sign(\sum_{i \epsilon SV} \alpha_i y_i (\boldsymbol{x_i} \cdot \boldsymbol{x}) + b) \end{aligned} \tag{2.21}$$

SVM linear dapat memisahkan data menggunakan sebuah *hyperplane*, sedangkan kenyataannya banyak data yang tidak dapat dipisahkan secara linear oleh sebuah *hyperplane*. Untuk menangani hal tersebut, diperkenalkan fungsi kernel pada penerapan SVM nonlinear.

2.1.3 Fungsi Kernel

Fungsi Kernel digunakan sebagai parameter dalam SVM untuk memetakan data yang tidak dapat dipisahkan secara linear ke ruang berdimensi lebih tinggi

sehingga data tersebut lebih mudah dipisahkan secara linear. SVM nonlinear melakukan pemetaan data dari ruang vektor awal $\{\boldsymbol{x_i} \in \mathbb{R}^Q\}$ yang berdimensi Q ke ruang baru $\{\boldsymbol{x_i} \in \mathbb{R}^R\}$ yang berdimensi R dengan $Q < R$. Pemetaan tersebut dapat dinotasikan sebagai fungsi transformasi $\varphi(\boldsymbol{x})$ (Nugroho, 2007).



Sumber: (Nugroho, 2007)
Gambar 2.2 Memetakan Data ke Ruang Dimensi Tinggi

Menurut Pratiwi & Setyawan (2021), vektor data ($\boldsymbol{x}$) yang sulit dipisahkan di ruang asli (*input space*) dipindahkan ke dalam ruang baru berdimensi lebih tinggi (*feature space*) menggunakan fungsi pemetaan

$$\varphi\big(\phi: \boldsymbol{x} \rightarrow \phi(\boldsymbol{x})\big) \tag{2.22}$$

Fungsi pemetaan $\varphi(\boldsymbol{x})$ memiliki proses yang rumit sehingga sulit dihitung. Oleh karena itu, perhitungan *dot product* dapat menggantikan karena mampu dihitung di *input space* maupun *feature space*. Berdasarkan teori *mercer*, perhitungan *dot product* dapat diformulasikan dengan fungsi kernel atau *kernel trick* di mana data seolah-olah sudah ditransformasikan oleh fungsi pemetaan φ. Fungsi kernel membuat perhitungan menjadi lebih sederhana karena tidak perlu menghitung seluruh koordinat dalam ruang dimensi baru (Jabbar dkk., 2021). Fungsi kernel dapat dirumuskan pada Persamaan (2.23).

$$K(\boldsymbol{x_i}, \boldsymbol{x_j}) = \phi(\boldsymbol{x_i}) \cdot \phi(\boldsymbol{x_j}) \tag{2.23}$$

sehingga pada persamaan (2.20) dapat diformulasikan menjadi,

$$h(\boldsymbol{x}) = sign(\sum_{i=1, \boldsymbol{x_i} \in SV} \alpha_i y_i \phi(\boldsymbol{x_i}) \cdot \phi(\boldsymbol{x_j}) + b)$$

$$= sign(\sum_{i=1, \boldsymbol{x_i} \in SV} \alpha_i y_i K(\boldsymbol{x_i}, \boldsymbol{x_j}) + b) \tag{2.24}$$

Secara umum, fungsi kernel terdiri dari kernel linear, kernel polinomial, kernel *Radial Basis Function* (RBF), dan kernel *sigmoid* (Han dkk., 2012).

1. Kernel Linear

   Kernel linear adalah fungsi kernel paling sederhana karena hanya digunakan pada ruang asli tanpa transformasi ke ruang baru. Umumnya, kernel linear digunakan pada SVM linear. Perhitungan kernel linear dapat dirumuskan sebagai berikut:

   $$K(\boldsymbol{x_i}, \boldsymbol{x_j}) = \boldsymbol{x_i} \cdot \boldsymbol{x_j} \tag{2.25}$$

2. Kernel Polinomial

   Kernel polinomial adalah pengembangan dari kernel linear. Kernel polinomial menghitung *dot product* antara dua vektor, kemudian dimasukkan ke dalam derajat lebih dari satu. Kernel polinomial dapat dirumuskan sebagai berikut:

   $$K(\boldsymbol{x_i}, \boldsymbol{x_j}) = (\boldsymbol{x_i} \cdot \boldsymbol{x_j} + c)^h \tag{2.26}$$

   Parameter yang digunakan yaitu h (*degree*).

3. *Kernel Radial Basis Function* (RBF)

   Kernel RBF menggunakan jarak *Euclidean* untuk menghitung jarak atau kemiripan antara dua data di ruang aslinya. Nilai kernel RBF bergantung pada jarak antara dua data. Kernel RBF dapat dirumuskan sebagai berikut:

$$K(\boldsymbol{x_i}, \boldsymbol{x_j}) = \exp\left(-\frac{\|\boldsymbol{x_i} - \boldsymbol{x_j}\|^2}{2\sigma^2}\right), \sigma \in \mathbb{R}^+ \tag{2.27}$$

   di mana $\|\boldsymbol{x_i} - \boldsymbol{x_j}\|$ adalah jarak *Euclidean* dan parameter yang digunakan yaitu σ.

4. Kernel *Sigmoid*

   Fungsi kernel *sigmoid* dapat dirumuskan sebagai berikut:

$$K(\boldsymbol{x_i}, \boldsymbol{x_j}) = \tanh(k\boldsymbol{x_i} \cdot \boldsymbol{x_j} - \delta) \tag{2.28}$$

   Parameter yang digunakan yaitu k dan δ.

2.1.4 *Adaptive Boosting*

Adaptive Boosting (AdaBoost) merupakan algoritma *machine learning* yang menggabungkan model klasifikasi sederhana (*weak learner*) menjadi model klasifikasi yang kuat untuk meningkatkan nilai akurasi pada model klasifikasi (Pison dkk., 2023). Algoritma AdaBoost diperkenalkan oleh Freund dan Schapire. Konsep algoritma tersebut adalah menyesuaikan bobot pada data latih di setiap iterasi. Penyesuaian dilakukan dengan memberikan bobot yang lebih besar pada data yang sulit diklasifikasikan agar model lebih fokus pada data tersebut di iterasi berikutnya. Proses ini akan memudahkan untuk mengklasifikasikan data yang sulit atau salah klasifikasi (Schapire, 2013).

Menurut Schapire (2013), tahap awal AdaBoost yaitu memberikan bobot yang sama pada setiap data latih yang dinotasikan $(\boldsymbol{x_1}, y_1), (\boldsymbol{x_2}, y_2), \dots, (\boldsymbol{x_n}, y_n)$ dengan $y_i \in \{-1,1\}$. Pemberian bobot dapat dirumuskan sebagai berikut:

$$D_1(i) = \frac{1}{n}, \qquad i = 1,2, \dots, n \tag{2.29}$$

dengan $D_1(i)$ adalah bobot awal atau iterasi pertama pada data ke-i.

AdaBoost melatih *weak learner* dengan menggunakan distribusi bobot di setiap iterasi ke-t. Proses ini menghasilkan model klasifikasi sederhana, yaitu $h_t(\boldsymbol{x})$ yang dipetakan ke dalam label kelas $\{-1,1\}$. Kemudian, hitung tingkat *weighted error* (ε_t) untuk memilih model sederhana dengan tingkat *weighted error* paling kecil. *Weighted error* (ε_t) dapat dinyatakan dengan persamaan sebagai berikut (Schapire, 2013):

$$\varepsilon_t = P_{i \sim D_1}[h_t(\boldsymbol{x_i}) \neq y_i] \tag{2.30}$$

dengan P_i adalah probabilitas data ke-i sesuai distribusi bobot pada iterasi pertama. Apabila hipotesis menghasilkan akurasi yang lebih baik dari prediksi, maka bobot α_t dapat ditambahkan pada hipotesis dengan rumus:

$$\alpha_t = \frac{1}{2} \ln\left(\frac{1-\varepsilon_t}{\varepsilon_t}\right) \tag{2.31}$$

Berdasarkan Freund & Schapire (1995), nilai α_t ditentukan oleh tingkat kesalahan ε_t. Nilai ε_t harus kurang dari 0.5 atau $\varepsilon_t \leq 0.5 - \gamma$ dengan γ adalah kelebihan dari performa nilai *weak classifier* di mana $\gamma > 0$. Apabila $\varepsilon_t < 0.5$, maka bobot akan bernilai positif sehingga model sederhana memberikan kontribusi yang baik. Apabila $\varepsilon_t = 0.5$, maka bobot $\alpha_t = 0$ berarti model tidak memberikan kontribusi terhadap model klasifikasi. Sebaliknya, apabila $\varepsilon_t > 0.5$,

maka α_t akan bernilai negatif sehingga model tidak layak digunakan dalam proses boosting.

Bobot data akan diperbarui tergantung hasil klasifikasi. Data yang salah klasifikasi akan diberikan bobot yang lebih besar, sedangkan data yang sudah benar diklasifikasi akan dikurangi nilai bobotnya. Pembaruan bobot pada iterasi selanjutnya dapat diformulasikan (Schapire, 2013):

$$D_{t+1}(i) = \frac{D_t(i)\exp\left(-\alpha_t y_i h_t(\boldsymbol{x_i})\right)}{Z_t} \tag{2.32}$$

dengan $D_t(i)$ adalah bobot pada iterasi ke-t dan Z_t adalah faktor normalisasi yang digunakan untuk memastikan $\sum_{i=1}^{n} D_{t+1}(i) = 1$ yaitu ketika jumlah semua bobot sampai data ke-n pada iterasi ke-$(t+1)$ adalah satu. Nilai Z_t diperoleh dari penjumlahah seluruh pembaruan bobot pada iterasi ke-t yang dapat dinyatakan pada Persamaan (2.33).

$$Z_t = \sum_{i=1}^{n} D_t(i)\exp\left(-\alpha_t y_i h_t(\boldsymbol{x_i})\right) \tag{2.33}$$

Setelah penyesuaian bobot, *weak learner* akan lebih fokus pada data yang salah diklasifikasikan untuk iterasi selanjutnya. Setelah mencapai iterasi maksimum (T), model klasifikasi gabungan yang diperoleh adalah (Schapire, 2013):

$$G(\boldsymbol{x}) = sign\left(\sum_{t=1}^{T'} \alpha_t h_t(\boldsymbol{x})\right) \tag{2.34}$$

Di mana T' adalah banyak iterasi yang memenuhi kriteria AdaBoost dengan $\varepsilon_t < 0.5$. Iterasi AdaBoost dilakukan sebanyak 10 iterasi untuk memperbaiki kesalahan klasifikasi tanpa menambah kompleksitas model dan menghindari *overfitting* akibat penggunaan iterasi yang banyak pada data yang terbatas.

2.1.5 K-*Fold Cross Validation*

K-*Fold Cross Validation* adalah teknik validasi untuk memastikan model yang dihasilkan dapat berfungsi dengan baik dan stabil pada data baru yang belum terlihat sebelumnya (Lumumba dkk., 2024). Proses validasi dilakukan dengan membagi dataset menjadi k lipatan dengan ukuran sama besar. Kemudian, pilih satu lipatan sebagai data uji dan pilih lipatan lainnya $(k-1)$ sebagai data latih. Gunakan data uji untuk menghitung *error* model dan ulangi proses tersebut sebanyak k kali sehingga tiap lipatan akan memiliki kesempatan untuk dipilih menjadi data uji. Dalam teknik ini, nilai k yang sering digunakan yaitu 5 atau 10 (James dkk., 2023). Perhitungan nilai *error* pada K-*Fold Cross Validation* dapat dituliskan sebagai (Lumumba dkk., 2024):

$$K-fold\ CV\ error = \frac{1}{k}\sum_{i=1}^{k} E_i \tag{2.35}$$

dengan,

$$E_i = \frac{1}{|D_i|}\sum_{j=1}^{|D_i|} L\left(y_{ij}, \hat{y}_{ij}\right) \tag{2.36}$$

di mana:

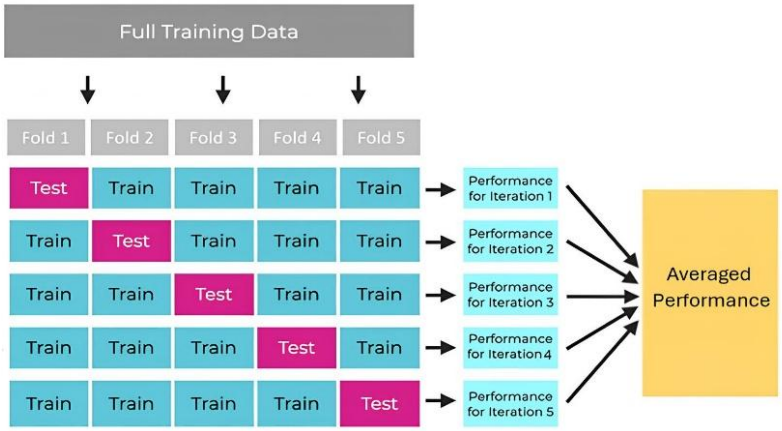
E_i : Nilai *error* pada lipatan ke-i,

D_i : Lipatan ke-i,

y_{ij} : Nilai asli dari data uji,

$\hat{y}_{ij}$: Nilai prediksi model pada data,

L : Fungsi kerugian.



Sumber:(Caprioli dkk., 2025)

Gambar 2.3 Ilustrasi *K-fold Cross Validation* dengan $K = 5$

2.1.6 *Confusion Matrix*

Confusion matrix adalah teknik evaluasi kinerja sebuah model klasifikasi yang berupa tabel atau matriks berbentuk persegi berukuran $N \times N$, di mana N merupakan jumlah kelas. Tabel tersebut berfungsi sebagai perbandingan antara nilai sebenarnya dengan nilai prediksi untuk mengetahui nilai keakuratan dari model yang dihasilkan (Sathyanarayanan & Tantri, 2024). Tabel *confusion matrix* dapat dilihat pada Tabel 2.1.

Tabel 2.1 *Confusion Matrix*

Predicted	***Actual***	
	True	***False***
True	TP (*True Positive*)	FP (*False Positive*)
False	TN (*True Negative*)	FN (*False Negative*)

Confusion matrix memiliki empat komponen utama, yaitu (Sathyanarayanan & Tantri, 2024):

1. *True Positive* (TP), ketika data positif dapat diklasifikasikan benar.
2. *True Negative* (TN), ketika data negatif dapat diklasifikasikan benar.

3. *False Positive* (FP), terjadi *error* tipe 1 di mana data negatif salah diklasifikasikan.
4. *False Negative* (FN), terjadi *error* tipe 2 di mana data positif salah diklasifikasikan.

Confusion matrix dapat digunakan juga untuk melakukan berbagai perhitungan dalam metrik evaluasi, yaitu (Sathyanarayanan & Tantri, 2024):

1. Akurasi, digunakan untuk mengetahui tingkat ketepatan model klasifikasi, yaitu persentase jumlah data yang berhasil diklasifikasikan dengan benar.

$$akurasi = \frac{TP + TN}{TP + FP + FN + TN} \tag{2.37}$$

2. *Recall*, digunakan untuk mengetahui kemampuan model dalam mendeteksi kelas positif, yaitu persentase kelas positif aktual yang berhasil diklasifikasikan dengan benar sebagai kelas positif.

$$recall = \frac{TP}{TP + FN} \tag{2.38}$$

3. Presisi, digunakan untuk mengukur ketepatan model dalam memprediksi kelas positif, yaitu persentase jumlah data yang diprediksi sebagai kelas positif dan benar termasuk kelas positif.

$$presisi = \frac{TP}{TP + FP} \tag{2.39}$$

4. F1-*score*, digunakan untuk menilai keseimbangan antara presisi dan *recall*, yaitu rata-rata dari kedua metrik tersebut untuk menggambarkan kinerja model klasifikasi yang lebih akurat.

$$F1 - score = 2 \times \frac{recall \times presisi}{recall + presisi} \tag{2.40}$$

2.2 Kanker Paru-Paru

Kanker paru-paru adalah penyakit yang muncul akibat adanya pertumbuhan sel atau jaringan abnormal yang tidak terkendali di paru-paru. Sel atau jaringan tersebut tumbuh dan menyebar dengan cepat ke organ tubuh lainnya. Faktor utama yang menyebabkan hal ini yaitu gaya hidup, seperti kebiasaan merokok, kurangnya asupan buah dan sayuran, dan jarang berolahraga (Wahdah dkk., 2024).

Kanker paru-paru terjadi karena sel atau jaringan normal di paru-paru mengalami kerusakan pada genetik sehingga sel tumbuh dengan ganas menjadi sel yang abnormal. Sel abnormal ini akan membentuk tumor yang dapat merusak jaringan kanker paru-paru dan sekitarnya. Ketika sel kanker sudah menyebar ke organ tubuh lainnya, sel tersebut akan membentuk tumor baru. Proses penyebaran ini dikenal dengan *metasis* (Asiyah dkk., 2025).

Faktor risiko merupakan paparan yang menyebabkan seseorang dapat terdiagnosis kanker paru-paru. Paparan ini biasanya berasal dari perilaku sehari-hari, keadaan lingkungan sekitar, maupun faktor biologis. Kanker paru-paru disebabkan oleh banyak faktor. Namun, faktor yang paling utama seseorang dapat terdiagnosis kanker paru-paru yaitu kebiasaan merokok (Buana & Harahap, 2022). Pada tahun 2020, tercatat bahwa kanker yang paling banyak diderita oleh laki-laki di Indonesia adalah kanker paru-paru dengan prevelensi perokok lebih dari 70%. Terdapat 34.783 kasus baru kanker paru-paru pada pria. Sedangkan, tercatat sebanyak 30.843 atau sekitar 13.2% mengalami kasus kematian. Dengan demikian, kanker paru-paru menjadi penyebab kasus kematian terbanyak di Indonesia (Asmara dkk., 2023).

Menurut Kementerian Kesehatan (2023), beberapa faktor risiko yang menjadi pemicu kanker paru-paru adalah sebagai berikut:

1. Usia

   Angka diagnosis kanker paru-paru termasuk rendah pada seseorang berusia di bawah 40 tahun. Namun, angka diagnosis terus meningkat ketika seseorang sudah memasuki usia lanjut hingga kelompok usia 70 tahun (Kementerian Kesehatan, 2023). Umumnya, kanker paru-paru akan mudah menyerang seseorang yang memasuki usia 55 tahun (Putri dkk., 2024).

2. Jenis kelamin

   Kanker paru-paru merupakan penyakit yang paling sering diderita oleh laki-laki dengan angka kematian kanker paru-paru 1/3 dari seluruh kematian akibat kanker dan menjadi kanker tersering kelima pada perempuan (Kementerian Kesehatan, 2023).

3. Status merokok

   Faktor risiko yang paling umum menyebabkan kanker paru-paru adalah merokok. Merokok menyebabkan seseorang terdiagnosis kanker paru-paru dengan kasus 80% pada laki-laki dan 50% pada perempuan (Kementerian Kesehatan, 2023).

4. *Komorbiditas*

   Komorbiditas adalah penyakit penyerta yang dialami oleh pasien kanker paru-paru yang dapat memengaruhi perjalanan penyakit. Penyakit penyerta ini yaitu penyakit paru obstruktif kronik (PPOK), hipertensi, diabetes, tuberkulosis, atau penyakit fibrosis paru idiopatik (Kementerian Kesehatan, 2023). Riwayat

penyakit tuberkulosis adalah salah satu penyakit penyerta yang paling sering ditemukan pada pasien kanker paru-paru mencapai 7.3% (Nastiti dkk., 2023).

5. Riwayat Keluarga

   Riwayat keluarga berpengaruh signifikan dengan risiko terdiagnosis kanker paru-paru. Faktor risiko ini dapat berupa faktor keturunan atau kebiasaan merokok dalam keluarga. Faktor ini juga akan terlihat lebih jelas pada non-perokok dan berisiko lebih tinggi pada usia muda di bawah 55 tahun (Gao dkk., 2009).

6. Stadium Kanker Paru-Paru

   Stadium kanker terdiri dari beberapa tingkat keparahan, yaitu:

   a. Stadium I, sel kanker hanya terdapat di paru-paru dan belum menyebar ke organ tubuh lainnya.
   b. Stadium II dan III, sel kanker sudah menyebar ke bagian kelenjar limfe di sekitar paru-paru tanpa adanya penyebaran jauh dari organ paru-paru.
   c. Stadium IV, sel kanker sudah menyebar ke organ tubuh lain di luar paru-paru.

   Pasien yang baru terdiagnosis kanker paru-paru ketika sudah memasuki stadium lanjut (stadium III dan IV) hanya memiliki peluang 1-5% untuk bertahan hidup. Sedangkan, pasien yang terdiagnosis lebih awal akan memiliki peluang hidup lebih besar 40-50% untuk bertahan hidup hingga lima tahun (Rejeki & Pratiwi, 2020).

Banyak penderita kanker paru-paru yang baru terdiagnosis ketika sudah mencapai stadium akhir di mana sel kanker sudah tumbuh dan menyebar dengan cepat ke organ tubuh lainnya. Hal tersebut menyebabkan kanker sulit disembuhkan.

Saat ini, pemeriksaan diagnosis lebih awal memiliki banyak risiko, seperti mudah terpapar radiasi yang tinggi. Oleh karena itu, upaya mencegah terkena kanker paru-paru sangat penting untuk meningkatkan kesadaran masyarakat terhadap faktor-faktor risikonya sehingga meningkatkan peluang deteksi penyakit lebih awal (Alfarisa dkk., 2023).

2.3 Kajian Integrasi Topik dengan Al-Quran/Hadits

Al-Qur'an tidak secara spesifik menjelaskan terkait kanker paru-paru. Namun, perintah untuk menjauhkan diri dari hal-hal yang membinasakan dapat ditemui di dalam Al-Qur'an surat Al-Baqarah ayat 195. Ayat tersebut menegaskan bahwa Allah SWT. memperingati manusia agar menghindari perilaku maupun kebiasaan yang dapat merusak kesehatan, salah satunya kebiasaan merokok. Merokok menjadi faktor utama penyebab kanker paru-paru yang membahayakan tubuh serta bertentangan dengan prinsip dalam islam. Ayat ini juga menegaskan pentingnya tanggung jawab manusia dalam menjaga amanah Allah SWT. dalam menjaga kesehatan diri sendiri.

Pengobatan kanker paru-paru dilakukan bergantung pada beberapa faktor, seperti jenis kanker, lokasi sel kanker, maupun kondisi pasien. Pengobatan ini memiliki berbagai pendekatan medis, yaitu operasi, kemoterapi, radioterapi, dan lain sebagainya. Tujuan dari pengobatan adalah untuk membunuh serta menghambat pertumbuhan sel agar tidak menyebar ke organ tubuh yang lain (Rusdi dkk., 2023). Meskipun pasien yang memasuki stadium lanjut hanya memiliki peluang hidup yang lebih rendah, pengobatan tetap penting dilakukan untuk memperlambat perkembangan sel, mengurangi gejala, dan membantu

meningkatkan kualitas hidup pasien. Sebagaimana Rasulullah SAW bersabda dalam HR. Muslim 4084 yang artinya:

"*Setiap penyakit mempunyai obatnya, jika obatnya mengenai penyakit, maka sembuhlah dengan izin Allah SWT*" (Muslim, 1998).

Hadits di atas mengandung arti bahwa setiap penyakit pasti memiliki obatnya karena Allah SWT. menciptakan segala sesuatu secara berpasangan, termasuk penyakit dengan obatnya. Namun, dari banyaknya obat, Allah SWT. adalah penyembuh sejati dan obat hanyalah perantara. Oleh karena itu, dalam melakukan pengobatan, seseorang tidak cukup dengan pengobatan medis. Pengobatan juga harus disertai dengan ikhtiar, berdoa, dan bertawakkal kepada Allah SWT. (Badrudin, 2020).

Upaya pencegahan dan pengobatan kanker paru-paru tidak hanya menjadi tanggung jawab tenaga medis saja, namun juga tanggung jawab seseorang dalam menjaga amanah tubuh dari Allah SWT. Menjauhi diri dari kebiasaan yang dapat merusak kesehatan dan melakukan pengobatan sesuai aturan medis menjadi bentuk ikhtiar seseorang dalam menjaga dirinya.

2.4 Kajian Topik dengan Teori Pendukung

Penelitian ini mengkaji bagaimana penerapan *Support Vector Machine* (SVM) yang dioptimalkan dengan *Adaptive Boosting* (AdaBoost) untuk menangani klasifikasi pasien kanker paru-paru berdasarkan faktor risikonya. Kanker paru-paru menjadi penyakit dengan angka kematian yang tinggi sehingga pentingnya upaya melakukan deteksi lebih awal melalui informasi faktor-faktor risiko kanker paru-paru. Salah satu upaya penelitian yang dapat dilakukan yaitu mengklasifikasikan pasien kanker paru-paru berdasarkan faktor risiko yang dialami untuk memberikan

informasi mengenai stadium kanker yang diderita. Informasi tersebut dapat membantu tenaga medis serta meningkatkan kesadaran bagi masyarakat untuk menghindari faktor-faktor yang memicu pernyakit tersebut.

Banyak penelitian terdahulu yang mengkaji penerapan algoritma AdaBoost-SVM pada kasus klasifikasi dalam bidang medis. Penelitian oleh Listiana & Muslim (2017) yang menerapkan algoritma AdaBoost-SVM pada kasus klasifikasi diagnosa penyakit ginjal kronis. Pada penelitian ini, spesifikasi model SVM menggunakan SVM kernel linear dan parameter C yaitu 10 dengan iterasi optimum berhenti pada iterasi ketiga di mana pada iterasi tersebut mampu meningkatkan nilai akurasi pada model. Nilai akurasi yang diperoleh dari gabungan algoritma tersebut mencapai 99.5% yang kemudian akan dibandingkan dengan algoritma SVM tunggal yang hanya menghasilkan nilai akurasi 62.5%. Berdasarkan penelitian ini menujukkan bahwa AdaBoost mampu mengoptimalkan performa SVM mencapai 37%.

Penelitian serupa oleh Adyalam dkk. (2018) juga menerapkan gabungan AdaBoost-SVM pada kasus klasifikasi penyakit *osteoarthritis*. Penelitian ini dilakukan menggunakan kernel RBF dengan melakukan percobaan pada setiap proporsi pembagian data latih dan data uji dan menghasilkan pembagian 80%:20% memiliki nilai akurasi paling tinggi. Proses pada penelitian ini dilakukan sebanyak 10 iterasi. Penelitian ini menghasilkan nilai akurasi sebesar 85.714% di mana akurasi tersebut lebih tinggi dibanding akurasi pada SVM tunggal yaitu hanya sebesar 68.572%.

Terdapat penelitian pada kasus klasifikasi kanker paru-paru oleh Sinaga dkk. (2022) menggunakan gabungan algoritma AdaBoost dengan *Random Forest*. Pada

penelitian ini, model diuji dengan dua proporsi pembagian data yaitu 80%:20% dan 70%:30%. Hasil yang diperoleh pada pembagian 80%:20% menghasilkan nilai akurasi tertinggi yaitu ketika jumlah pohon yang dibentuk sebanyak 147 pohon dengan kedalaman maksimum dibatasi hingga 5 dan parameter kontrol yaitu 95. Proses tersebut dilakukan sebanyak 154 iterasi pada boosting. Hasil penelitian ini memberikan nilai akurasi mencapai 95.4% dibanding nilai akurasi pada model *Random Forest* tunggal yaitu 93.20%.

Penelitiaan ini akan menerapkan algoritma AdaBoost-SVM pada kasus klasifikasi pasien kanker paru-paru. Berdasarkan penelitian oleh Wang (2012) menerapkan algoritma SVM yang dioptimalkan dengan AdaBoost melalui beberapa tahapan. Proses di mulai dari memberikan bobot yang sama pada setiap data latih dengan Persamaan (2.29). Selanjutnya, melatih model SVM pada Persamaan (2.21) dengan bobot awal, kernel linear, dan parameter C yang digunakan. Setelah model SVM dilatih, hitung *weighted error* (ε_t) menggunakan Persamaan (2.30) di mana jika hasil *weighted error* kurang dari 0.5, gunakan nilai tersebut untuk memperoleh nilai bobot pengaruh pada model SVM menggunakan Persamaan (2.31). Selanjutnya, perbarui bobot berdasarkan hasil klasifikasi menggunakan Persamaan (2.32). Data yang salah diklasifikasikan akan diberikan bobot yang lebih besar supaya algoritma dapat lebih fokus pada data tersebut pada iterasi selanjutnya, sedangkan data yang benar diklasifikasikan akan diberikan bobot yang lebih kecil. Proses tersebut dilakukan secara berulang hingga mencapai iterasi optimum. Kemudian, model akhir dibentuk dengan menggabungkan beberapa model SVM yang telah dilatih menggunakan Persamaan (2.34).

BAB III

METODE PENELITIAN

3.1 Jenis Penelitian

Jenis pendekatan yang digunakan dalam penelitian ini adalah pendekatan kuantitatif. Data yang digunakan yaitu data sekunder berupa data rekam medis pasien kanker paru-paru di RSUD Karsa Husada Batu tahun 2020-2025. Data tersebut akan dianalisis klasifikasi menggunakan metode *Support Vector Machine* yang akan dioptimalkan dengan *Adaptive Boosting*.

3.2 Data dan Sumber Data

Data yang digunakan dalam penelitian ini adalah data sekunder berupa data pasien kanker paru-paru pada tahun 2020-2025. Data diperoleh dari rekam medis RSUD Karsa Husada Batu berupa laporan riwayat diagnosa dari masing-masing pasien. Data mencakup lima variabel *independent* (X) yang menjadi faktor risiko kanker paru-paru berdasarkan penelitian oleh Harahap dkk. (2017) dan satu variabel *dependent* (Y). Kategorisasi variabel-variabel yang digunakan pada penelitian ini terdapat pada Tabel 3.1:

Tabel 3.1 Variabel Penelitian

No.	Variabel		Keterangan
1.	Usia	X_1	Usia pasien saat menjalani pemeriksaan atau pengobatan di RSUD Karsa Husada
2.	Jenis Kelamin	X_2	(1) Laki-laki (2) Perempuan
3.	Status Merokok	X_3	(1) Merokok (2) Tidak merokok

No.	Variabel		Keterangan
4.	Komorbiditas	X_4	(1) Ya, mempunyai komorbiditas (2) Tidak mempunyai komorbiditas
5.	Riwayat Keluarga	X_5	(1) Ya, memiliki riwayat keluarga (2) Tidak memiliki riwayat keluarga
6.	Stadium Kanker Paru-Paru	Y	(1) Stadium III (-1) Stadium IV

3.3 Teknik Pengumpulan Data

Teknik pengumpulan data yang dilakukan pada penelitian ini yaitu menggunakan teknik dokumenter yaitu mengumpulkan data dengan cara melihat, mencatat, serta menganalisis dokumen yang tersedia. Dokumen yang digunakan yaitu rekam medis pasien kanker paru-paru di RSUD Karsa Husada Batu tahun 2020-2025 berupa laporan riwayat diagnosa dari masing-masing pasien yang dapat dilihat pada Lampiran 1.

3.4 Tahapan Penelitian

Adapun langkah-langkah yang dilakukan penelitian ini sebagai berikut:

1. Mengumpulkan data rekam medis pasien kanker paru-paru di RSUD Karsa Husada Batu.
2. Menginput data penelitian yang akan diklasifikasikan.
3. *Preprocessing* data.
    a. Memberikan label atau kategori pada variabel yang ingin diklasifikasikan, misalnya 1 untuk stadium III dan -1 untuk stadium IV.
    b. Melakukan pembersihan data kategorik dari *missing value* menggunakan nilai modus.

c. Melakukan standardisasi data variabel usia menggunakan Persamaan (2.7).

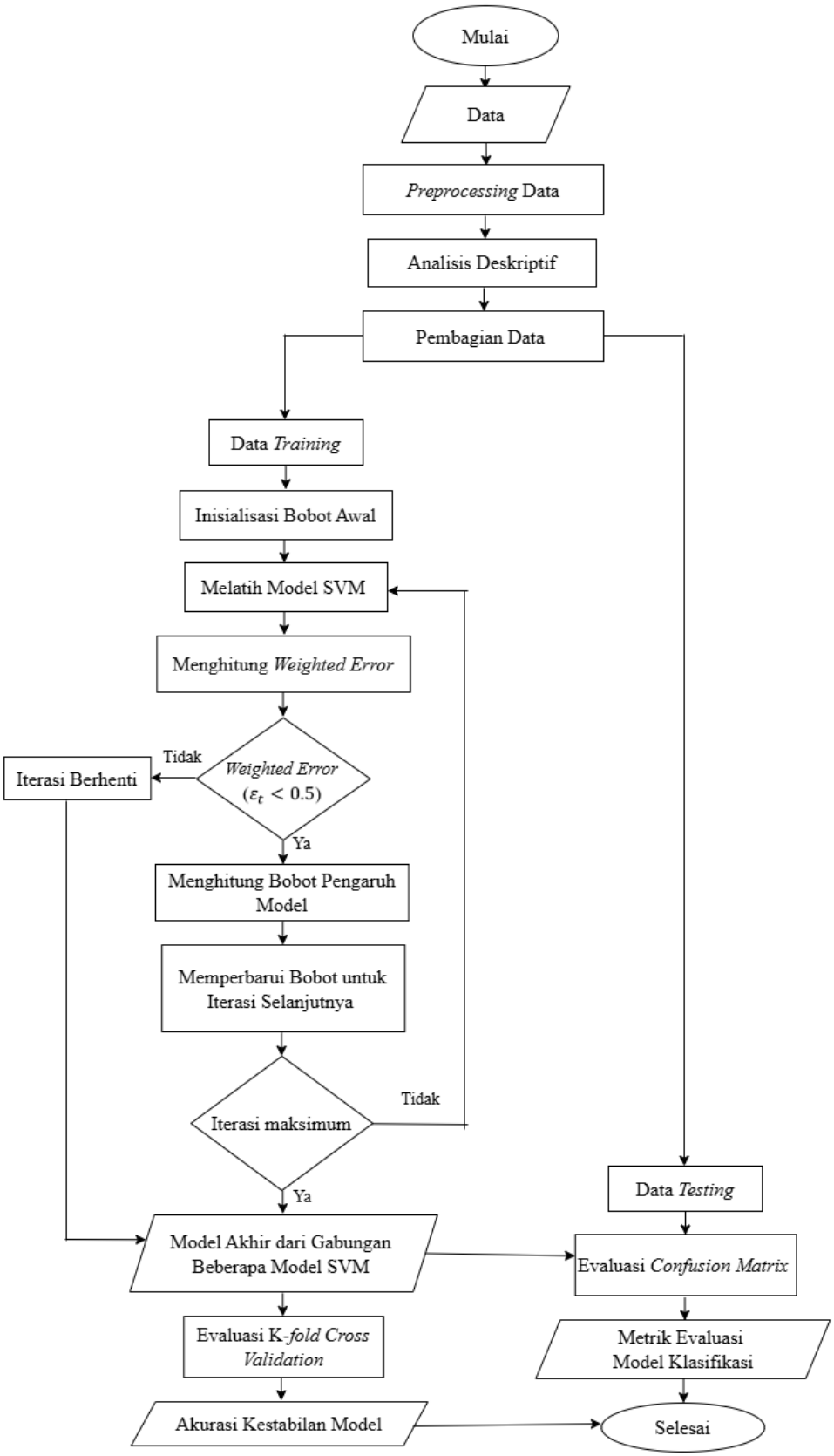
4. Melakukan analisis deskriptif.
5. Melakukan pembagian data menjadi data *training* dan data *testing* dengan rasio 90%:10%. Rasio tersebut dipilih berdasarkan hasil pengujian yang menunjukkan nilai akurasi terbaik melalui proses *trial and error*.
6. Membangun model klasifikasi menggunakan algoritma AdaBoost-SVM. Langkah-langkah dalam menerapkan algoritma tersebut sebagai berikut:
   a. Inisialisasi bobot awal pada data *training*. Setiap data *training* akan diberikan bobot yang sama. Perhitungan bobot awal menggunakan Persamaan (2.29).
   b. Melatih model SVM dengan kernel linear menggunakan Persamaan (2.21) dengan bobot awal yang diperoleh. Parameter kernel yang digunakan yaitu $C = 0.01; 0.1; 1; 10; 100$.
   c. Menghitung *weighted error* menggunakan Persamaan (2.30).
      - Apabila $\varepsilon_t > 0.5$, model yang dihasilkan tidak layak.
      - Apabila $\varepsilon_t = 0.5$, model yang dihasilkan tidak memberikan pengaruh pada proses *boosting*.
      - Apabila $\varepsilon_t < 0.5$, model yang dihasilkan berpengaruh baik sehingga perlu menghitung bobot pengaruh model dengan Persamaan (2.31).
   d. Melakukan pembaruan bobot pada data *training* menggunakan Persamaan (2.32) dengan menaikkan nilai bobot pada data yang salah diklasifikasikan dan menurunkan nilai bobot pada data yang benar diklasifikasikan.

e. Lakukan langkah b sampai d secara berulang hingga mencapai iterasi maksimum. Banyaknya iterasi maksimum yang digunakan adalah 10 iterasi.

f. Membentuk model klasifikasi akhir dengan menggabungkan beberapa model SVM yang memenuhi kriteria $\varepsilon_t < 0.5$ menggunakan Persamaan (2.34).

7. Melakukan evaluasi model dengan K-*Fold Cross Validation* menggunakan Persamaan (2.35) dan menghitung metrik performa pada data uji dengan *confusion matrix* menggunakan Persamaan (2.37), (2.38), (2.39), dan (2.40).
8. Melakukan interpretasi dari hasil evaluasi model klasifikasi. Hasil evaluasi akan dianalisis untuk menilai tingkat akurasi model klasifikasi berdasarkan metrik evaluasi.

3.5 Diagram Alir Penelitian



Gambar 3.1 Diagram Alir Penelitian

BAB IV

HASIL DAN PEMBAHASAN

4.1 *Preprocessing* Data

Preprocessing data merupakan tahap awal sebelum proses pemodelan klasifikasi. Tahap ini dilakukan untuk memastikan data yang akan digunakan telah bersih dari data yang hilang dan berada pada skala data yang sesuai sehingga algoritma dapat memproses dengan baik. Dalam penelitian ini, *preprocessing data* yang dilakukan meliputi penanganan *missing value* dan standardisasi data.

Data yang digunakan yaitu data pasien kanker paru-paru tahun 2020 sampai 2025 yang diambil dari rekam medis RSUD Karsa Husada Batu dengan 132 data. Variabel *independent* yang digunakan yaitu Usia (X_1), Jenis Kelamin (X_2), Status Merokok (X_3), Komorbiditas (X_4), dan Riwayat Keluarga (X_5), sedangkan variabel *dependent* yang digunakan yaitu Stadium Kanker Paru-Paru (Y) yang meliputi stadium III dan stadium IV.

Tahap pertama *preprocessing* data yaitu penanganan *missing value* yang dilakukan untuk membersihkan data dari nilai yang hilang. Berdasarkan karakteristik data pasien kanker paru-paru yang digunakan, terdapat *missing value* pada beberapa variabel, yaitu Stadium Kanker Paru-Paru (Y) sebesar 42.42% atau 56 data yang hilang, Status Merokok (X_3) sebesar 28.03% atau 37 data yang hilang, Komorbiditas (X_4) sebesar 1.52% atau 2 data yang hilang, dan Riwayat Keluarga (X_5) sebesar 14.39% atau 19 data yang hilang.

Missing value pada data medis merupakan permasalahan yang umum terjadi dalam penelitian klinis. Kondisi dapat disebabkan oleh berbagai faktor, seperti

pasien yang datang sudah dalam kondisi kritis, keterbatasan biaya dan waktu, serta adanya risiko tertentu bagi pasien apabila harus menjalani pengujian tambahan (Kusuma, 2015). Proporsi *missing value* yang cukup besar dapat mempengaruhi hasil analisis. Oleh karena itu, diperlukan teknik penanganan *missing value* yang tepat agar data dapat digunakan dengan baik.

Pengujian mekanisme *missing value* dilakukan dengan membandingkan kelompok data *missing* dan data tidak *missing* menggunakan variabel indikator, yaitu bernilai 1 jika data *missing* dan bernilai 0 jika data tidak *missing*. Selanjutnya, dilakukan pengujian hubungan atau perbedaan terhadap variabel lain yang teramati sebagai variabel pembanding. Pengujian pada variabel kategorik dilakukan menggunakan uji *Chi-square* atau uji *Fisher's Exact*, sedangkan pengujian pada variabel numerik dilakukan menggunakan uji *Mann Whitney*. Hasil pengujian untuk identifikasi *missing value* pada variabel kategorik menggunakan uji *Chi-Square* atau uji *Fisher's Exact* dapat dilihat pada Tabel 4.1.

Tabel 4.1 Hasil Uji *Chi-Square* dan Uji *Fisher's Exact*

Variabel *missing*	Variabel pembanding	Uji Statistik	χ^2_{hitung}	χ^2_{tabel}	$p-value$	Keputusan
X_3	X_2	*Chi-Square*	0.237	3.841	0.626	Tidak terdapat hubungan
X_4	X_2	*Fisher's Exact*	–	–	1.000	Tidak terdapat hubungan
X_5	X_2	*Chi-Square*	0.420	3.841	0.517	Tidak terdapat hubungan
Y	X_2	*Chi-Square*	0.742	3.841	0.389	Tidak terdapat hubungan

Berdasarkan Tabel 4.1, pada variabel X_3 diperoleh nilai χ^2_{hitung} sebesar 0.237 dengan $p-value$ sebesar 0.626, variabel X_5 diperoleh nilai χ^2_{hitung} sebesar 0.420 dengan $p-value$ sebesar 0.517, dan variabel Y diperoleh nilai χ^2_{hitung} sebesar 0.742 dengan $p-value$ sebesar 0.389. Hasil pengujian menunjukkan bahwa

seluruh variabel memiliki nilai $p - value > 0.05$. Sementara itu, pada variabel X_4 terhadap variabel pembanding X_2 digunakan uji *Fisher's Exact* karena terdapat frekuensi harapan kurang dari 5. Hasil pengujian memperoleh nilai $p - value$ sebesar 1.000, di mana $p - value > 0.05$. Berdasarkan hasil kedua pengujian tersebut menunjukkan terima H_0 sehingga dapat disimpulkan bahwa tidak terdapat hubungan yang signifikan antara indikator *missing value* dengan variabel pembanding.

Hasil pengujian untuk identifikasi *missing value* pada variabel numerik menggunakan uji *Mann-Whitney* dapat dilihat pada Tabel 4.2.

Tabel 4.2 Hasil Uji *Mann-Whitney*

Variabel *missing*	Variabel pembanding	Uji Statistik	$\|Z\|$	$Z_{1/2\alpha}$	$p - value$	Keputusan
X_3	X_1	*Mann-Whitney*	1.654	1.96	0.098	Tidak terdapat perbedaan
X_4	X_1	*Mann-Whitney*	0.494	1.96	0.628	Tidak terdapat perbedaan
X_5	X_1	*Mann-Whitney*	1.248	1.96	0.213	Tidak terdapat perbedaan
Y	X_1	*Mann-Whitney*	1.105	1.96	0.270	Tidak terdapat perbedaan

Statistik uji *Mann-Whitney* dilakukan menggunakan nilai Z karena ukuran sampel yang digunakan lebih dari 20. Berdasarkan Tabel 4.2, pada variabel X_3 diperoleh nilai $|Z|$ sebesar 1.654 dengan $p - value$ sebesar 0.098, variabel X_4 diperoleh nilai $|Z|$ sebesar 0.494 dengan $p - value$ sebesar 0.628, variabel X_5 diperoleh nilai $|Z|$ sebesar 1.248 dengan $p - value$ sebesar 0.213, dan variabel Y diperoleh nilai $|Z|$ sebesar 1.105 dengan $p - value$ sebesar 0.270. Hasil pengujian menunjukkan bahwa seluruh variabel memiliki nilai $p - value > 0.05$ sehingga terima H_0. Dengan demikian, dapat disimpulkan bahwa tidak terdapat perbedaan signifikan antara variabel indikator *missing value* dengan variabel pembanding.

Berdasarkan hasil pengujian *Chi-Square*, *Fisher's Exact,* dan *Mann-Whitney* menunjukkan bahwa *missing value* yang terjadi tidak memiliki hubungan maupun perbedaan signifikan terhadap variabel lain. Oleh karena itu, mekanisme *missing value* dapat diindikasikan sebagai *Missing Completely At Random* (MCAR) yaitu kondisi ketika data yang hilang terjadi secara acak tanpa adanya hubungan dengan variabel lain maupun variabel itu sendiri. Oleh karena itu, metode sederhana masih dapat digunakan untuk mengisi data yang hilang (Widyawati dkk., 2025). Salah satu metode imputasi sederhana yang dapat digunakan pada data kategorik adalah imputasi modus. Berdasarkan penelitian Memon dkk. (2023), imputasi modus menunjukkan performa yang cukup baik pada data medis kategorik meskipun proporsi *missing value* cukup besar, yaitu sebesar 30%, 40%, dan 50%. Metode ini dilakukan dengan menggantikan nilai yang hilang menggunakan kategori yang paling sering muncul pada masing-masing variabel. Setelah proses imputasi dilakukan, data yang digunakan telah bersih dari *missing value*.

Tahap kedua yaitu standardisasi data yang dilakukan menggunakan metode *z-score* untuk menyamakan skala variabel karena algoritma sensitif terhadap perbedaan skala. Berikut contoh perhitungan standardisasi data pada variabel usia menggunakan Persamaan (2.7).

$$Z_{11} = \frac{\mathrm{X}_{11} - \overline{\mathrm{X}}_1}{s_1}$$

$$= \frac{25 - 60.51}{11.41} = -3.11$$

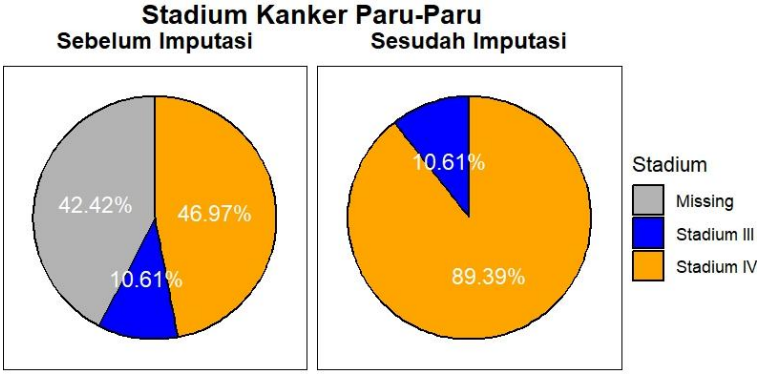
Hasil standardisasi data variabel usia dapat dilihat pada Tabel 4.3 dan Lampiran 2.

Tabel 4.3 Standardisasi Data

No	X_{1i}	Z_i
1	25	-3.11
2	42	-1.62
3	72	1.01
4	57	-0.31
5	61	0.04
6	78	1.53
⋮	⋮	⋮
128	49	-1.01
129	74	1.18
130	33	-2.41
131	69	0.74
132	47	-1.18

4.2 Analisis Deskriptif

Analisis deskriptif dilakukan guna melihat distribusi dari karakteristik data pasien kanker paru-paru. Analisis ini bertujuan untuk menggambarkan kondisi data sebelum dilakukan proses imputasi untuk melihat distribusi awal dari data pasien dan kondisi data sesudah dilakukan *preprocessing* data untuk melihat pengaruh penanganan *missing value* terhadap distribusi data. Terdapat enam variabel yang digunakan, yaitu Usia (X_1), Jenis Kelamin (X_2), Status Merokok (X_3), Komorbiditas (X_4), Riwayat Keluarga (X_5), dan stadium kanker paru-paru (Y).



Gambar 4.1 *Pie Chart* Stadium Sebelum dan Sesudah Imputasi

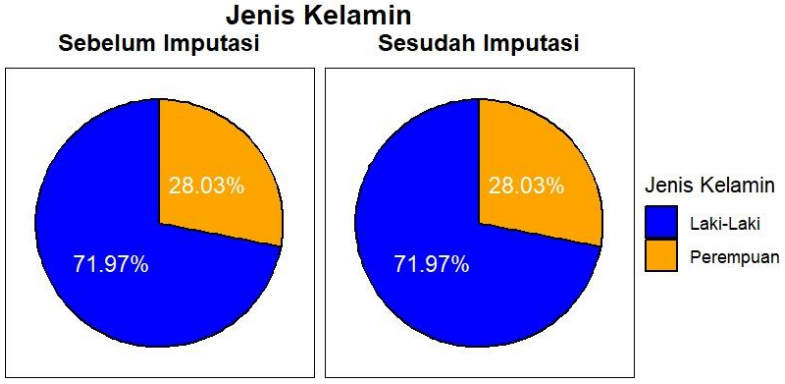
Berdasarkan Gambar 4.1 ditunjukkan persentase variabel stadium pasien kanker paru-paru sebelum dan sesudah dilakukan proses imputasi data. Sebelum

dilakukan imputasi, terdapat sebanyak 10.61% atau 14 pasien pada stadium III dan 46.97% atau 62 pasien pada stadium IV. Selain itu, terdapat 42.42% atau 56 data yang mengalami *missing value* sehingga perlu dilakukan proses imputasi data. Setelah dilakukan imputasi, *missing value* tidak lagi ditemukan. Persentase pasien stadium IV meningkat menjadi 89.39% atau 118 pasien, sedangkan pasien stadium III tetap 10.61% atau 14 pasien.

Tabel 4.4 Analisis Deskriptif Variabel Usia

Variabel	***Mean***	**Std. Deviasi**	**Minimum**	**Maksimum**
Usia	60.51	11.41	25	85

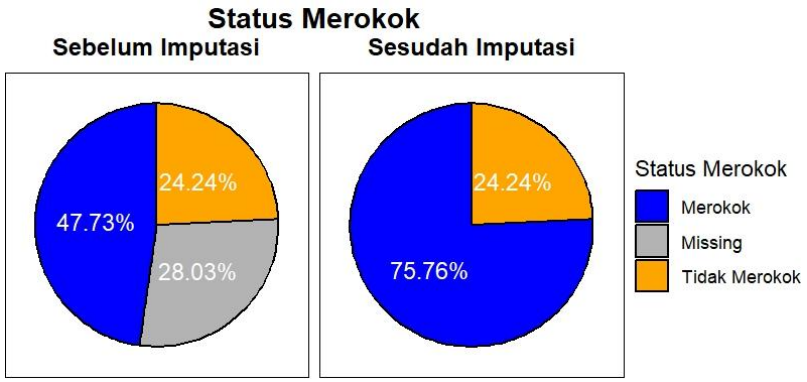
Variabel usia tidak mengalami *missing value* sehingga tidak memerlukan proses imputasi. Berdasarkan Tabel 4.4, diperoleh rata-rata usia pasien kanker paru-paru berada pada usia 60.51 tahun dengan standar deviasi sebesar 11.41. Hal ini menunjukkan bahwa pasien kanker paru-paru rata-rata pada kelompok usia lanjut. Usia termuda yaitu 25 tahun dan usia tertua pada usia 85 tahun.



Gambar 4.2 *Pie Chart* jenis Kelamin Sebelum dan Sesudah Imputasi

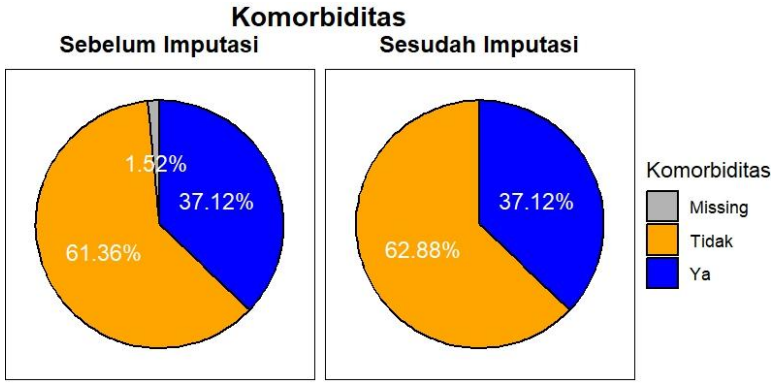
Variabel jenis kelamin tidak mengalami *missing value* sehingga tidak memerlukan proses imputasi. Berdasarkan Gambar 4.2 menunjukkan bahwa jenis kelamin laki-laki memiliki persentase sebesar 71.97% atau 95 pasien dan jenis kelamin perempuan memiliki persentase sebesar 28.03% atau 37 pasien. Hal ini

menunjukkan bahwa kejadian kanker paru-paru banyak ditemukan pada jenis kelamin laki-laki dibandingkan perempuan.



Gambar 4.3 *Pie Chart* Status Merokok Sebelum dan Sesudah Imputasi

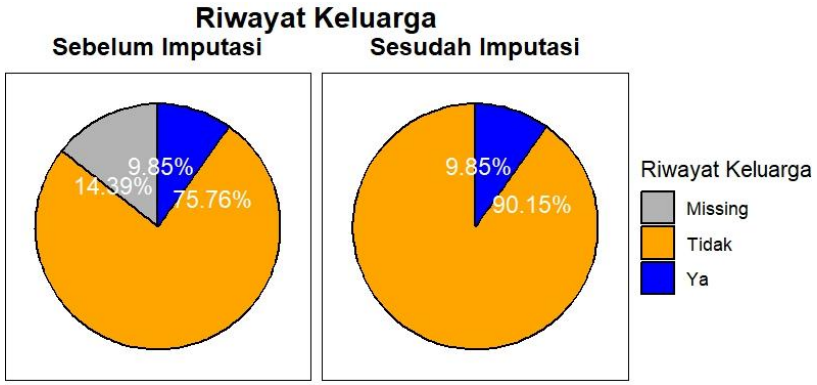
Berdasarkan Gambar 4.3 ditunjukkan persentase variabel status merokok sebelum dan sesudah dilakukan proses imputasi data. Sebelum dilakukan imputasi, terdapat 47.73% atau 63 pasien perokok dan 24.24% atau 32 pasien bukan perokok. Selain itu, terdapat 28.03% atau 37 data yang mengalami *missing value* sehingga perlu dilakukan proses imputasi data. Setelah dilakukan imputasi, *missing value* tidak lagi ditemukan. Persentase pasien perokok meningkat menjadi 75.76% atau 100 pasien, sedangkan persentase pasien bukan perokok tetap sebesar 24.24% atau 32 pasien.



Gambar 4.4 *Pie Chart* Komorbiditas Sebelum dan Sesudah Imputasi

Berdasarkan Gambar 4.4 ditunjukkan persentase variabel komorbiditas sebelum dan sesudah dilakukan proses imputasi data. Sebelum dilakukan imputasi,

terdapat 37.12% atau 49 pasien dengan komorbiditas dan 61.36% atau 81 pasien tanpa komorbiditas. Selain itu, terdapat 1.52% atau 2 data yang mengalami *missing value* sehingga perlu dilakukan proses imputasi data. Setelah dilakukan imputasi, *missing value* tidak lagi ditemukan. Persentase pasien tanpa komorbiditas sedikit meningkat menjadi 62.88% atau 83 pasien, sedangkan persentase pasien dengan komorbiditas tetap sebesar 37.12% atau 49 pasien.



Gambar 4.5 *Pie Chart* Riwayat Keluarga Sebelum dan Sesudah Imputasi

Berdasarkan Gambar 4.5 ditunjukkan persentase variabel riwayat keluarga sebelum dan sesudah dilakukan proses imputasi data. Sebelum dilakukan imputasi, terdapat 9.85% atau 13 pasien dengan riwayat keluarga dan 75.76% atau 100 pasien tanpa riwayat keluarga. Selain itu, terdapat 14.39% atau 19 data yang mengalami *missing value* sehingga perlu dilakukan proses imputasi data. Setelah dilakukan imputasi, *missing value* tidak lagi ditemukan. Persentase pasien tanpa riwayat keluarga meningkat menjadi 90.15% atau 119 pasien, sedangkan persentase pasien dengan riwayat keluarga tetap sebesar 9.85% atau 13 pasien.

4.3 Pembagian Data *Training* dan Data *Testing*

Tahap awal proses klasifikasi AdaBoost-SVM dilakukan dengan membagi data secara acak menjadi dua bagian, yaitu data *training* dan data *testing*. Pemilihan

persentase data *training* dan data *testing* dilakukan melalui beberapa percobaan dengan berbagai kombinasi persentase. Data *training* digunakan untuk membentuk model AdaBoost-SVM, sedangkan data *testing* digunakan untuk mengevaluasi ketepatan model klasifikasi yang diperoleh dari data *training*. Berikut beberapa kombinasi persentase yang digunakan dapat dilihat pada Tabel 4.5

Tabel 4.5 Metrik Evaluasi Perbandingan Data *Training* dan Data *Testing*

Data *Training*	**Data *Testing***	**Akurasi**	***Recall***	**Presisi**
90%	**10%**	**76.92%**	**50%**	**33.33%**
85%	15%	80%	0%	0%
80%	20%	61.54%	40%	22.22%
75%	25%	60.61%	40%	16.67%
70%	30%	80%	0%	0%

Berdasarkan Tabel 4.5, terlihat bahwa pada persentase 85%:15% dan 70%:30% menghasilkan nilai akurasi tertinggi yaitu sebesar 80%. Namun, kombinasi persentase tersebut menghasilkan model yang cenderung hanya memprediksi kelas negatif dan gagal mengidentifikasi kelas positif yang ditunjukkan oleh nilai *recall* dan presisi sebesar 0%. Oleh karena itu, kombinasi persentase yang digunakan pada penelitian ini adalah 90% atau 119 data *training* dan 10% atau 13 data *testing*. Pada proporsi data *training*, terdapat 107 pasien stadium IV dan 12 pasien stadium III. Sedangkan pada proporsi data *testing*, terdapat 11 pasien stadium IV dan 2 pasien stadium III.

4.4 Klasifikasi dengan Metode AdaBoost-SVM

Proses klasifikasi AdaBoost-SVM dilakukan menggunakan bantuan *software* RStudio. Sebelum proses pemodelan, data kategorik seperti Jenis Kelamin (X_2), Status Merokok (X_3), Komorbiditas (X_4), Riwayat Keluarga (X_5), dan Stadium Kanker Paru-Paru (Y) terlebih dahulu ditransformasikan ke dalam bentuk numerik

agar dapat diproses oleh algoritma SVM. Selain itu, variabel numerik seperti Usia (X_1) dilakukan standardisasi untuk menyamakan skala data sehingga semua variabel dapat memberikan kontribusi yang seimbang saat digunakan oleh algoritma.

4.4.1 Pelatihan Model Dasar SVM

Pelatihan model AdaBoost-SVM dilakukan menggunakan data yang telah melalui tahap *preprocessing* serta pembagian data dengan proporsi 90% data *training* (119 data) dan 10% data *testing* (13 data). Model AdaBoost-SVM dibentuk menggunakan kernel linear berdasarkan data *training*. Pada penelitian ini, SVM berperan sebagai *base learner* pada algoritma AdaBoost. Proses boosting dilakukan sebanyak 10 iterasi untuk memperbaiki kesalahan klasifikasi serta menghindari risiko *overfitting* akibat penggunaan iterasi yang terlalu banyak pada data yang terbatas. Pada setiap iterasi, model SVM akan dilatih menggunakan bobot data yang diperbarui berdasarkan kesalahan hasil klasifikasi pada iterasi sebelumnya. Data yang salah diklasifikasikan akan diberikan bobot yang lebih besar, sedangkan data yang benar diklasifikasikan diberikan bobot yang lebih kecil.

Tahap pertama yang dilakukan adalah inisialisasi bobot awal di mana bobot awal bernilai sama untuk semua data *training* sebanyak 119 data. Bobot awal dapat dituliskan pada persamaan (2.29) sehingga diperoleh perhitungan bobot awal sebagai berikut:

$$D_1(i) = \frac{1}{n}$$

$$= \frac{1}{119}$$

$$= 0.0084$$

Bobot awal pada iterasi pertama yaitu sebesar 0.0084. Pemberian bobot awal yang sama pada seluruh data *training* menunjukkan bahwa setiap data *training* dianggap memiliki tingkat kepentingan yang sama dalam proses pembentukan model klasifikasi pada iterasi pertama. Selanjutnya, bobot tersebut digunakan dalam pelatihan model klasifikasi.

Tahap berikutnya yaitu melatih SVM dengan memanfaatkan nilai bobot yang diperoleh pada setiap iterasi. Penelitian ini menggunakan kernel linear dengan parameter *cost* (C) yaitu 0.01; 0.1; 1; 10; 100. Pemilihan parameter tersebut berdasarkan pada penelitian Kurnia dkk. (2018). Penentuan parameter terbaik dilakukan melalui proses *trial and error*. Model AdaBoost-SVM yang terbentuk dari data *training* kemudian diuji menggunakan data *testing* untuk memperoleh nilai evaluasi model pada setiap parameter kernel linear yang dapat dilihat pada Tabel 4.6.

Tabel 4.6 Hasil Metrik Evaluasi Parameter Kernel Linear

Parameter	**Akurasi**	***Recall***	**Presisi**
0.01	84.62%	0%	0%
0.1	84.62%	0%	0%
1	69.23%	50%	25%
10	**76.92%**	**50%**	**33.33%**
100	**76.92%**	**50%**	**33.33%**

Berdasarkan Tabel 4.6, diperoleh nilai evaluasi dari model AdaBoost-SVM pada berbagai kombinasi parameter. Terlihat bahwa parameter $C = 0.01$ dan $C = 0.1$ menghasilkan akurasi tertinggi sebesar 84.62%. Namun, meskipun parameter

tersebut menghasilkan nilai akurasi yang cukup tinggi, model yang dihasilkan tidak dapat memprediksi kelas positif pada data yang bersifat tidak seimbang (*imbalanced*) yang dapat dilihat melalui *confusion matrix* pada Lampiran 3 dan Lampiran 4. Hal tersebut ditunjukkan oleh nilai *recall* dan presisi sebesar 0%. Oleh karena itu, penelitian ini tidak hanya mempertimbangkan nilai akurasi tertinggi, tetapi juga memperhatikan model yang dapat mengklasifikasikan kedua kelas melalui metrik evaluasi *recall* dan presisi. Berdasarkan pertimbangan tersebut, parameter terbaik untuk membangun model AdaBoost-SVM dengan kernel linear adalah parameter $C = 10$ dan $C = 100$ yang menghasilkan nilai akurasi sebesar 76.92% serta nilai *recall* dan presisi yang sama. Karena kedua parameter menghasilkan performa yang sama, maka dalam penelitian ini dipilih parameter yang lebih sederhana yaitu $C = 10$.

Berdasarkan parameter terbaik yang digunakan, dihitung matriks kernel yang diperoleh dari hasil perkalian dua vektor dan memiliki ukuran $n \times n$ di mana n sebanyak data *training*. Berikut perhitungan untuk elemen matriks kernel 119×119 menggunakan Persamaan (2.25) dengan contoh perhitungan pada $K(\boldsymbol{x}_1, \boldsymbol{x}_1)$:

$$K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \boldsymbol{x}_i \cdot \boldsymbol{x}_j$$

$$\begin{aligned} K(\boldsymbol{x}_1, \boldsymbol{x}_1) &= x_{11} \cdot x_{11} + x_{12} \cdot x_{12} + x_{13} \cdot x_{13} + x_{14} \cdot x_{14} + x_{15} \cdot x_{15} \\ &= (-1.183 \cdot -1.183) + (2 \cdot 2) + (1 \cdot 1) + (2 \cdot 2) + (2 \cdot 2) \\ &= 14.4 \end{aligned}$$

Perhitungan di atas dapat diterapkan pada seluruh data *training* sehingga diperoleh matriks kernel sebagai berikut:

$$K = \begin{bmatrix} 14.4 & 10.4 & \cdots & 12.9 \\ 10.4 & 8.4 & \cdots & 8.9 \\ \vdots & \vdots & \ddots & \vdots \\ 12.9 & 8.9 & \cdots & 13 \end{bmatrix}$$

Selanjutnya, proses optimasi untuk memperoleh nilai $\boldsymbol{w}$ dan b pada iterasi pertama dilakukan menggunakan bantuan *software*. Nilai $\boldsymbol{w}$ yang dihasilkan menggunakan Persamaan (2.16) dapat dilihat pada Tabel 4.7.

Tabel 4.7 Nilai Vektor Bobot $\boldsymbol{w}$

$\boldsymbol{w_i}$	**Nilai**
$\boldsymbol{w_1}$	0
$\boldsymbol{w_2}$	0.0009
$\boldsymbol{w_3}$	0.0005
$\boldsymbol{w_4}$	−0.0009
$\boldsymbol{w_5}$	−0.0001

Adapun nilai b yang diperoleh menggunakan Persamaan (2.14) adalah -1.0005. Dengan nilai $\boldsymbol{w}$ dan b yang telah diperoleh, model SVM dapat dibentuk berdasarkan Persamaan (2.21).

$$\begin{aligned} h_t(\boldsymbol{x_i}) &= sign(\boldsymbol{w} \cdot \boldsymbol{x_i} + b) \\ h_1(\boldsymbol{x_i}) &= sign((0 \cdot x_{i1}) + (0.0009 \cdot x_{i2}) + (0.0005 \cdot x_{i3}) + (-0.0009 \cdot x_{i4}) \\ &\quad + (-0.0001 \cdot x_{i5}) - 1.0005) \\ &= sign(0.0009x_{i2} + 0.0005x_{i3} - 0.0009x_{i4} - 0.0001x_{i5} \\ &\quad - 1.0005) \end{aligned} \tag{4.1}$$

Berdasarkan model SVM yang telah terbentuk pada Persamaan (4.1), proses klasifikasi dapat dilakukan dengan perhitungan sebagai berikut:

$$\begin{aligned} h_1(\boldsymbol{x_1}) &= sign(0.0009(2) + 0.0005(1) - 0.0009(2) - 0.0001(2) - 1.0005) \\ &= sign(0.0018 + 0.0005 - 0.0018 - 0.0002 - 1.0005) \\ &= sign(-1.0002) \end{aligned}$$

Hasil perhitungan menunjukkan bahwa data pertama diklasifikasikan sebagai kelas negatif. Proses perhitungan diterapkan pada seluruh data *training* sehingga diperoleh 12 data yang salah diklasifikasikan, yaitu data ke-15, 28, 56, 60, 81, 82, 93, 98, 99, 112, 114, 117. Selanjutnya, hasil klasifikasi tersebut digunakan dalam proses boosting pada AdaBoost untuk menghitung *weighted error* dan pembaruan bobot pada iterasi selanjutnya.

4.4.2 Perhitungan *Weighted Error*

Perhitungan *weighted error* (ε_t) digunakan untuk mengukur tingkat kesalahan klasifikasi model pada setiap iterasi AdaBoost. Nilai ε_t diperoleh dari jumlah bobot data yang salah diklasifikasikan. Berdasarkan model pada iterasi pertama, terdapat 12 data yang salah diklasifikasikan sehingga perhitungan ε_t dapat dilakukan menggunakan Persamaan (2.30) sebagai berikut:

$$\varepsilon_1 = P_{i\sim D_1}[h_t(\boldsymbol{x_i}) \neq y_i]$$

$$= P_{i\sim D_1}[h_1(\boldsymbol{x_{15}}) \neq y_{15}] + P_{i\sim D_1}[h_1(\boldsymbol{x_{28}}) \neq y_{28}] + \cdots$$

$$+ P_{i\sim D_1}[h_1(\boldsymbol{x_{114}}) \neq y_{114}] + P_{i\sim D_1}[h_1(\boldsymbol{x_{117}}) \neq y_{117}]$$

$$= 0.0084 + 0.0084 + \cdots + 0.0084 + 0.0084$$

$$= 0.1008$$

Berdasarkan hasil perhitungan *weighted error*, diperoleh $\varepsilon_1 = 0.1008$ di mana $\varepsilon_1 < 0.5$. Nilai tersebut menunjukkan bahwa model pada iterasi pertama masih memenuhi kriteria dalam pembentukan model akhir AdaBoost-SVM. Jika nilai *weighted error* melebihi 0.5, maka model memiliki tingkat kesalahan yang tinggi sehingga model tidak memberikan kontribusi yang baik untuk pembentukan model akhir.

4.4.3 Perhitungan Bobot Pengaruh Model

Bobot pengaruh α_t digunakan untuk mengukur seberapa besar kontribusi model sederhana dalam pembentukan model akhir AdaBoost-SVM. Berdasarkan hasil perhitungan *weighted error* pada iterasi pertama, nilai bobot pengaruh model dapat dihitung menggunakan Persamaan (2.31) sebagai berikut:

$$\alpha_1 = \frac{1}{2}\ln\left(\frac{1-0.1008}{0.1008}\right)$$

$$= \frac{1}{2}\ln\left(\frac{0.8992}{0.1008}\right)$$

$$= 1.094$$

Berdasarkan hasil perhitungan, α_1 memperoleh nilai sebesar 1.094. Nilai tersebut menunjukkan bahwa model SVM pada iterasi pertama memberikan kontribusi dalam pembentukan model akhir AdaBoost-SVM. Semakin besar nilai α_t, maka semakin besar kontribusi model sederhana dalam pembentukan model klasifikasi akhir. Selanjutnya, nilai α_1 digunakan dalam proses pembaruan bobot pada iterasi berikutnya.

4.4.4 Pembaruan Bobot Data

Pembaruan bobot dilakukan berdasarkan hasil klasifikasi pada iterasi sebelumnya. Data yang salah diklasifikasikan diberikan bobot yang lebih besar, sedangkan data yang benar diklasifikasikan diberikan bobot yang lebih kecil. Pembaruan bobot dilakukan menggunakan Persamaan (2.32) dengan contoh perhitungan pada data ke-1 sebagai data yang benar diklasifikasikan dan data ke-15 sebagai data yang salah diklasifikasikan.

Perhitungan pembaruan bobot pada data yang benar diklasifikasikan:

$$D_2(1) = \frac{D_1(1)\exp\left(-\alpha_1 y_1 h_1(\boldsymbol{x_1})\right)}{Z_t}$$

dengan,

$$Z_t = \left(D_1(1)\exp\left(\left(-1.094(-1)(-1)\right)\right)\right) + \cdots + \left(D_1(117)\exp\left(\left(-1.094(1)(-1)\right)\right)\right) + \left(D_1(118)\exp\left(\left(-1.094(-1)(-1)\right)\right)\right) + \left(D_1(119)\exp\left(\left(-1.094(-1)(-1)\right)\right)\right)$$

$$= 0.0084\exp(-1.094) + \cdots + 0.0084\exp(1.094) + 0.0084\exp(-1.094) + 0.0084\exp(-1.094)$$

$$= 0.0028 + \cdots 0.025 + 0.0028 + 0.028$$

$$= 0.5996$$

Sehingga,

$$D_2(1) = \frac{D_1(1)\exp\left(\left(-1.094(-1)(-1)\right)\right)}{0.5996}$$

$$= \frac{0.0028}{0.5996} = 0.0047$$

Perhitungan pembaruan bobot pada data yang salah diklasifikasikan:

$$D_2(15) = \frac{D_1(15)\exp\left(-\alpha_1 y_1 h_1(\boldsymbol{x_{15}})\right)}{Z_t}$$

$$= \frac{D_1(15)\exp\left(\left(-1.094(1)(-1)\right)\right)}{0.5996}$$

$$= \frac{0.025}{0.5996} = 0.0417$$

Berdasarkan perhitungan pembaruan bobot, bobot data yang salah diklasifikasikan meningkat menjadi 0.0417 sehingga pada iterasi berikutnya kontribusi data tersebut menjadi lebih dominan dibandingkan data lainnya. Peningkatan bobot ini menunjukkan bahwa data tersebut merupakan data yang

sulit diklasifikasikan pada iterasi sebelumnya sehingga AdaBoost akan memberikan perhatian yang lebih besar untuk meminimalkan kesalahan pada iterasi selanjutnya. Proses boosting pada algoritma AdaBoost dilakukan secara iteratif hingga mencapai 10 iterasi. Hasil perhitungan pada setiap iterasi dapat dilihat pada Tabel 4.8.

Tabel 4.8 Hasil Perhitungan *Weighted Error* dan Bobot Pengaruh Model

Iterasi (t)	*Weighted Error* (ε_t)	Alpha (α_t)
1	0.1008	1.094
2	0.4108	0.1803
3	0.4479	0.1045
4	0.4309	0.1391
5	0.402	0.1985
6	0.4804	0.0392
7	0.4683	0.0635
8	0.4838	0.0323
9	0.4628	0.0746
10	0.5232	–

Berdasarkan Tabel 4.8, nilai *weighted error* (ε_t) pada iterasi ke-1 hingga iterasi ke-9 yang diperoleh kurang dari 0.5 sehingga model SVM yang terbentuk pada iterasi tersebut masih dapat digunakan dalam pembentukan model AdaBoost-SVM. Nilai *weighted error* (ε_t) terkecil terdapat pada iterasi pertama yaitu sebesar 0.1008 sehingga memberikan kontribusi terbesar yaitu sebesar 1.094. Pada iterasi ke-10, nilai *weighted error* (ε_t) yang diperoleh sebesar 0.5232 di mana $\varepsilon_t > 0.5$ sehingga model pada iterasi tersebut tidak digunakan dan proses boosting dihentikan. Dengan demikian, model klasifikasi AdaBoost-SVM dibentuk berdasarkan gabungan dari model sederhana yang dihasilkan pada iterasi ke-1 hingga iterasi ke-9.

4.5 Evaluasi Model

Model AdaBoost-SVM yang telah terbentuk dievaluasi untuk mengetahui performa model dalam mengklasifikasikan data. Evaluasi model dilakukan menggunakan K-*Fold Cross Validation* dan *Confusion Matrix.* Evaluasi menggunakan K-*Fold Cross Validation* digunakan untuk mengukur kestabilan model, sedangkan evaluasi *Confusion Matrix* digunakan untuk mengukur ketepatan model dalam mengklasifikasikan data.

4.5.1 Evaluasi Model dengan K-*Fold Cross Validation*

Evaluasi model AdaBoost-SVM menggunakan metode K-*Fold Cross Validation* dilakukan untuk menguji kestabilan model yang dihasilkan berdasarkan data *training*. Pada penelitian ini, digunakan jumlah *fold* (k) sebanyak 5 *fold* sehingga data *training* akan dibagi menjadi lima bagian. Pada setiap *fold*, empat bagian data digunakan sebagai data *training* dan satu bagian lainnya digunakan sebagai data *testing*. Model AdaBoost-SVM akan dilatih dan diuji masing-masing sebanyak lima kali sehingga seluruh bagian data akan mendapatkan giliran sebagai data *testing*. Hasil rata-rata akurasi yang diperoleh menggunakan 5-*Fold Cross Validation* ditunjukkan pada Tabel 4.9.

Tabel 4.9 Hasil Evaluasi Menggunakan 5-*Fold Cross Validation*

Fold	**Akurasi Fold**
$k = 1$	91.3%
$k = 2$	88%
$k = 3$	91.67%
$k = 4$	91.3%
$k = 5$	87.5%
Rata-rata	**89.96%**

Berdasarkan Tabel 4.9 ditunjukkan bahwa nilai rata-rata akurasi pada model AdaBoost-SVM menggunakan kernel linear dengan parameter $C = 10$ adalah 89.96%. Hasil akurasi tersebut menunjukkan bahwa model AdaBoost-SVM yang dihasilkan pada data *training* memiliki tingkat kestabilan yang baik dalam mengklasifikasikan pasien kanker paru-paru berdasarkan faktor risiko.

4.5.2 Evaluasi Model dengan *Confusion Matrix*

Evaluasi dengan *Confusion Matrix* dilakukan untuk mengetahui ketepatan klasifikasi model terhadap data baru yang tidak terlibat pada tahap pemodelan. Proses ini dilakukan menggunakan data *testing* dengan metrik evaluasi *confusion matrix*, yaitu akurasi, *recall*, presisi, dan F1-*Score*. Model yang diuji merupakan model terbaik yang diperoleh menggunakan kernel linear dengan parameter $C = 10$. Berdasarkan parameter tersebut, diperoleh tabel hasil *confusion matrix* yang dapat dilihat pada Tabel 4.10 sebagai berikut:

Tabel 4.10 *Confusion Matrix* AdaBoost-SVM

Confusion Matrix		Aktual	
		Stadium III	Stadium IV
Prediksi	Stadium III	1	2
	Stadium IV	1	9

Berdasarkan Tabel 4.10, diketahui sebanyak 1 pasien stadium III diklasifikasikan dengan benar sebagai pasien stadium III, sedangkan 1 pasien stadium III salah diklasifikasi sebagai pasien stadium IV. Selain itu, sebanyak 9 pasien stadium IV yang berhasil diklasifikasikan dengan benar sebagai pasien stadium IV, sedangkan 2 pasien stadium IV salah diklasifikasikan sebagai pasien stadium III. Berdasarkan hasil klasifikasi tersebut, diperoleh perhitungan metrik evaluasi berupa akurasi, presisi, *recall*, dan F1-*Score*.

$$akurasi = \frac{1+9}{1+1+2+9} \times 100\% = 76.92\%$$

$$recall = \frac{1}{1+1} \times 100\% = 50\%$$

$$presisi = \frac{1}{1+2} \times 100\% = 33.33\%$$

$$F1 - score = 2 \times \frac{50\% \times 33.33\%}{50\% + 33.33\%} \times 100\% = 40\%$$

Berdasarkan hasil perhitungan metrik evaluasi, diperoleh nilai akurasi sebesar 76.92% yang menunjukkan bahwa model AdaBoost-SVM mampu mengklasifikasikan sebagian besar data dengan benar. Nilai *recall* yang diperoleh sebesar 50% yang menunjukkan bahwa model mampu mengidentifikasi 50% atau setengah dari total pasien stadium III. Hal tersebut mengindikasikan bahwa masih terdapat sebagian pasien stadium III yang diklasifikasikan sebagai stadium IV karena distribusi data yang tidak seimbang (*imbalanced*) di mana pasien stadium IV merupakan kelas mayoritas dibandingkan pasien stadium III. Dalam kondisi tersebut, model cenderung lebih sering mempelajari pola kelas mayoritas sehingga mempengaruhi kemampuan dalam mengenali kelas minoritas.

Selain itu, nilai presisi yang diperoleh sebesar 33.33% yang menunjukkan bahwa dari seluruh data yang diprediksi sebagai stadium III, sebesar 33.33% yang merupakan stadium III sebenarnya. Kemudian, nilai F1-*Score* yang diperoleh sebesar 40% merupakan nilai keseimbangan antara nilai *recall* dan presisi. Hasil ini menunjukkan kemampuan model dalam mempertimbangkan kedua metrik tersebut pada kelas minoritas.

Kondisi ini berkaitan dengan mekanisme bobot (D_i) AdaBoost di mana bobot akan dinaikkan pada data yang salah diklasifikasikan dan diturunkan pada

data yang benar diklasifikasikan. Namun, karena kelas negatif merupakan kelas mayoritas, total bobot pada kelas tersebut dapat tetap lebih besar. Akibatnya, model cenderung lebih banyak mempelajari pola pada kelas mayoritas dibandingkan kelas minoritas. Dengan demikian, nilai akurasi yang diperoleh menunjukkan kemampuan model dalam mengklasifikasikan data secara keseluruhan, terutama pada kelas mayoritas.

4.6 Kajian Penelitian dalam Perspektif Islam

Kanker paru-paru merupakan penyakit yang mencakup seluruh keganasan yang berkembang di organ paru dan berasal dari jaringan paru itu sendiri (primer). Secara klinis, kanker paru-paru primer merupakan tumor ganas yang tumbuh dan berkembang dari jaringan epitel yang melapisi saluran penapasan, terutama bronkus. Penyakit ini berawal dari jaringan epitel bronkus yang mengalami perubahan sel tidak normal dan terus berkembang secara tidak terkendali sehingga membentuk tumor ganas. Pertumbuhan sel yang tidak terkendali tersebut dapat merusak jaringan paru dan sekitarnya. Oleh karena itu, kanker paru-paru primer sering disebut karsinoma bronkus atau *bronchogenic carcinoma* (Kementerian Kesehatan, 2023).

Berdasarkan hasil penelitian, model klasifikasi AdaBoost-SVM mampu mengklasifikasikan sebagian besar data pasien dengan benar sehingga dapat membantu proses identifikasi kondisi pasien berdasarkan faktor risiko yang dimiliki. Hasil tersebut menunjukkan bahwa pemanfaatan ilmu pengetahuan dan teknologi merupakan salah satu bentuk ikhtiar dalam mendukung proses diagnosis penyakit agar dapat dilakukan secara lebih cepat dan akurat. Dengan demikian,

penerapan metode AdaBoost-SVM diharapkan dapat membantu upaya identifikasi pasien sehingga dapat mendukung upaya deteksi dini kanker paru-paru.

Menurut pandangan islam, kesehatan menjadi salah satu nikmat yang harus dijaga dan dipelihara. Salah satu cara untuk menjaganya yaitu dengan tidak membiarkan diri terjerumus ke dalam kebinasaan. Sebagaimana yang dijelaskan dalam QS. Al-Baqarah ayat 195 yang menegaskan pentingnya menjaga diri dari kebinasaan dan tetap berbuat baik dalam menghadapi suatu permasalahan. Dalam konteks penelitian ini, ayat tersebut dapat dimaknai sebagai peringatan untuk menjaga tubuh kita dengan menghindari faktor-faktor yang dapat mengganggu kesehatan. Upaya tersebut dapat dilakukan melalui pemanfaatan ilmu pengetahuan untuk membantu mengidentifikasi kondisi pasien secara lebih cepat.

Ikhtiar yang dilakukan manusia juga sejalan dengan upaya melakukan berbagai pengobatan dan pencarian solusi atas setiap penyakit. Meskipun pasien pada stadium lanjut yang memiliki peluang hidup lebih rendah, pengobatan tetap perlu dilakukan sebagai usaha untuk memperlambat perkembangan sel kanker dan meningkatkan kualitas hidup pasien. Sebagaimana sabda Rasulullah SAW dalam HR. Muslim (1998) No. 4048 menyatakan bahwa setiap penyakit pasti memiliki obatnya. Meskipun demikian, dari banyaknya obat, hanya Allah SWT. yang menjadi penyembuh sejati dan obat hanyalah perantara. Oleh karena itu, penerapan metode AdaBoost-SVM diharapkan dapat menjadi salah satu bentuk dukungan dalam membantu proses deteksi dini kanker paru-paru berdasarkan faktor risiko pasien.

BAB V

PENUTUP

5.1 Kesimpulan

Berdasarkan rumusan masalah dan pembahasan pada penelitian ini, maka diperoleh kesimpulan sebagai berikut:

1. Hasil klasifikasi pasien kanker paru-paru berdasarkan faktor risiko telah berhasil dilakukan menggunakan algoritma *Adaptive Boosting* (AdaBoost)-*Support Vector Machine* (SVM). Model AdaBoost-SVM terbaik yang diperoleh yaitu menggunakan kernel linear dengan parameter *cost* $(C) = 10$. Hasil klasifikasi menunjukkan sebanyak 1 pasien stadium III diklasifikasikan benar sebagai pasien stadium III, 1 pasien stadium III salah diklasifikasikan sebagai pasien stadium IV, 9 pasien stadium IV diklasifikasikan benar sebagai pasien stadium IV, dan 2 pasien stadium IV salah diklasifikasikan sebagai pasien stadium III.
2. Tingkat akurasi dari performa model AdaBoost-SVM yang diperoleh menggunakan kernel linear dengan parameter $C = 10$ sebesar 76.92%. Nilai tersebut menunjukkan bahwa model mampu mengklasifikasikan sebagian besar data pasien kanker paru-paru dengan benar. Berdasarkan hasil metrik evaluasi lainnya, diperoleh nilai *recall* sebesar 50% yang menunjukkan bahwa model berhasil mengidentifikasi 50% atau setengah dari total pasien stadium III. Selain itu, nilai presisi diperoleh sebesar 33.33% yang menunjukkan bahwa dari seluruh data yang diprediksi sebagai pasien stadium III, sebesar 33.33% yang merupakan pasien stadium III sebenarnya. Kemudian, nilai F1-*Score* yang diperoleh sebesar 40% yang menunjukkan keseimbangan antara nilai

presisi dan *recall* dalam mengklasifikasikan pasien stadium III. Hasil tersebut dipengaruhi oleh kondisi data yang digunakan merupakan data tidak seimbang (*imbalanced)*.

5.2 Saran

Adapun beberapa hal yang dapat dijadikan acuan untuk penyempurnaan penelitian selanjutnya antara lain penggunaan dataset yang lebih besar untuk mendapatkan model klasifikasi yang lebih stabil. Penambahan dataset dapat berupa penambahan rentang periode kasus maupun penambahan fitur atau atribut dari faktor terjadinya kanker paru-paru. Selain itu, penelitian selanjutnya disarankan dapat menerapkan skema pembobotan lain untuk memberikan perhatian yang lebih besar pada kelas minoritas sehingga dapat meningkatkan performa model serta dapat menggunakan metrik evaluasi lainnya.

DAFTAR PUSTAKA

Abushark, Y. B., Hassan, S., & Khan, A. I. (2025). Optimized Adaboost Support Vector Machine-Based Encryption for Securing IoT-Cloud Healthcare Data. *Sensors*, *25*(3). https://doi.org/10.3390/s25030731

Adyalam, T. R., Rustam, Z., & Pandelaki, J. (2018). Classification of Osteoarthritis Disease Severity Using Adaboost Support Vector Machines. *Journal of Physics: Conference Series*, *1108*(1). https://doi.org/10.1088/1742-6596/1108/1/012062

Alfarisa, S., Mitra, E., & Wahyuni, S. (2023). Karakteristik Pasien Kanker Paru di RSUP Dr. M. Djamil Padang Tahun 2021. *Scientific Journal*, *2*(6), 141–149. https://doi.org/10.56260/sciena.v2i6.116

Almaidah, F., & Ambarwati, D. (2022). Perbandingan Gambaran Ct Scan Paru Perokok dan Non Perokok Pasien Kanker Paru. *Jurnal Kesehatan*, *VII*(Ii), 20–27.

Asiyah, S. N., Rachman, F., Syaiba, A., dkk. (2025). Peran Gizi Dalam Menangani Kanker Paru-Paru *: LITERATURE REVIEW The Role Of Nutrition In Managing Lung Cancer : Literature Review*. *32*.

Asmara, O. D., Tenda, E. D., Singh, G., dkk. (2023). Lung Cancer in Indonesia. *Journal of Thoracic Oncology*, *18*(9), 1134–1145. https://doi.org/10.1016/j.jtho.2023.06.010

Badrudin. (2020). Berobat Menurut Islam. *Jurnal Kependidikan Dan Keislaman*, *7*(1), 1–20.

Buana, I., & Harahap, D. A. (2022). Asbestos, Radon Dan Polusi Udara Sebagai Faktor Resiko Kanker Paru Pada Perempuan Bukan Perokok. *AVERROUS: Jurnal Kedokteran Dan Kesehatan Malikussaleh*, *8*(1), 1. https://doi.org/10.29103/averrous.v8i1.7088

Caprioli, L., Najlaoui, A., Campoli, F., dkk. (2025). Tennis Timing Assessment by a Machine Learning-Based Acoustic Detection System: A Pilot Study. *Journal of Functional Morphology and Kinesiology*, *10*(1). https://doi.org/10.3390/jfmk10010047

Cortes, C., & Vapnik, V. (1995). *Support-Vector Networks*. *297*, 273–297.

Freund, Y., & Schapire, R. E. (1995). A Decision-Theoretic Generalization Of On-Line Learning And An Application To Boosting. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *904*, 23–37. https://doi.org/10.1007/3-540-59119-2_166

Gao, Y., Goldstein, A. M., Consonni, D., dkk. (2009). Family History of Cancer and Non-Malignant Lung Diseases as Risk Factors For Lung Cancer. *Int J Cancer*, *125*(1), 146–152. https://doi.org/10.1002/ijc.24283

García, S., Luengo, J., & Herrera, F. (2015). Data Preprocessing in Data Mining. In *Intelligent Systems Reference Library* (Vol. 72). https://doi.org/10.1007/978-3-319-10247-4_8

Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques (3rd ed.)*. Morgan Kaufmann. https://doi.org/https://doi.org/10.1016/C2009-0-61819-5

Harahap, S. P., Sutandyo, N., Rumende, C. M., & Shatri, H. (2017). Perbandingan Rejimen Kemoterapi Cisplatin Etoposide dengan Cisplatin-Docetaxel dalam Hal Kesintasan 2 Tahun dan Progression-Free Survival Pasien Kanker Paru Stadium Lanjut Jenis Non-Small Cell. *Jurnal Penyakit Dalam Indonesia*, *3*(2), 67. https://doi.org/10.7454/jpdi.v3i2.11

Jabbar, M. A., Abraham, A., Dogan, O., dkk. (2021). Deep Learning in Biomedical and Health Informatics. In *Deep Learning in Biomedical and Health Informatics*. https://doi.org/10.1201/9781003161233

James, G., Witten, D., Hastie, T., dkk. (2023). An Introduction Statistical Machine Learning With Applications in Python. *Springer Texts in Statistics*, 425–472. https://www.statlearning.com/

Kemenag. (2024). *Qur'an Kemenag*. https://quran.kemenag.go.id/quran/per-ayat/surah/2?from=195&to=286

Kementerian Kesehatan, R. (2023). Panduan Penatalaksanaan Kanker Paru. *Komite Penanggulangan Kanker Nasional*, 10–11.

Kurnia, A. I., Furqon, M. T., & Rahayudi, B. (2018). *Klasifikasi Kualitas Susu Sapi Menggunakan Algoritme Support Vector Machine (SVM) (Studi Kasus : Perbandingan Fungsi Kernel Linier dan RBF Gaussian)*. *2*(11), 4453–4461.

Kusuma, D. H. (2015). Penanganan Missing Value Dalam Penyusunan Basis Kasus Pada Case-Based Reasoning. *I*(1).

Listiana, E., & Muslim, M. A. (2017). Penerapan Adaboost Untuk Klasifikasi Support Vector Machine Guna Meningkatkan Akurasi Pada Diagnosa Chronic Kidney Disease. *2015*, 875–881.

Lumumba, V. W., Kiprotich, D., Makena, N. G., & Kavita, M. D. (2024). *Comparative Analysis of Cross-Validation Techniques: LOOCV, K-folds Cross-Validation, and Repeated K-folds Cross-Validation in Machine Learning Models*. *13*(5), 127–137.

Mahendro, I., & Abimanto, D. (2023). Algoritma Support Vector Machine Untuk Menganalisa Kepuasan Mahasiswa Terhadap Pembelajaran E-Learning. *CV Pustaka Stimart AMNI*, *11*(1), 1–14. https://search.app/Kya2Fs5V4SAzm1cq7

Maulana, A., Sarwido, S., & Sucipto, A. (2025). Optimasi Algoritma Support Vector Machines(Svm) Menggunakan Adaptive Boosting(Adaboost) Untuk Meningkatkan Akurasi Prediksi Penyakit Brain Stroke. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *9*(5), 7614–7620. https://doi.org/10.36040/jati.v9i5.14779

Mchugh, M. L. (2013). *Lessons in Biostatistics The Chi-Square Test Of Independence*. *23*(2), 143–149.

Memon, S. M., Wamala, R., & Kabano, I. H. (2023). A Comparison Of Imputation Methods For Categorical Data. *Informatics in Medicine Unlocked*, *42*(September), 101382. https://doi.org/10.1016/j.imu.2023.101382

Muslim, A. H. A. H. (1998). Shohih Muslim. *Bait Al-Afkar*.

Nabeel, S. M., Bazai, S. U., Alasbali, N., dkk. (2024). Optimizing Lung Cancer Classification Through Hyperparameter Tuning. *Digital Health*, *10*. https://doi.org/10.1177/20552076241249661

Nastiti, N. P., Wulandari, L., Sulistiawati, S., dkk. (2023). Prevalence of Lung Cancer with a History of Tuberculosis. *Jurnal Respirasi*, *9*(2), 87–92. https://doi.org/10.20473/jr.v9-i.2.2023.87-92

Nicholson, J. S., Deboeck, P. R., & Howard, W. (2016). *Attrition in developmental psychology : A review of modern missing data reporting and practices*. https://doi.org/10.1177/0165025415618275

Nugroho, A. S. (2007). Pengantar Support Vector Machine. *Nagoya Jepang: Intitute of Technology*, 3.

Pison, N. A., Eka Suryana, M., & Irzal, M. (2023). *Prototype System Klasifikasi Genus Ikan Menggunakan Viola-Jones Feature Extraction dan Boosting Berbasis Decision Tree*. 43–48.

Pratiwi, N., & Setyawan, Y. (2021). Analisis Akurasi Dari Perbedaan Fungsi Kernel Dan Cost Pada Support Vector Machine Studi Kasus Klasifikasi Curah Hujan Di Jakarta. *Journal of Fundamental Mathematics and Applications (JFMA)*, *4*(2), 203–212. https://doi.org/10.14710/jfma.v4i2.11691

Putri, D. S. R., Wulanningrum, D. N., & Suryandari, D. (2024). Edukasi Kegawatdaruratan Kanker Paru pada Keluarga dalam Merawat Pasien di Rumah. *Jurnal Peduli Masyarakat*, *6*(2), 207–212.

Ranganathan, S., Nakai, K., & Schonbach, C. (2018). *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics*. Elsevier Science. https://books.google.co.id/books?id=rs51DwAAQBAJ

Rejeki, M., & Pratiwi, E. N. (2020). Diagnosis dan Prognosis Kanker Paru, Probabilitas Metastasis dan Upaya Prevensinya. *Proceeding of The URECOL*, *1*(1), 73–78.

Rusdi, N. K., Sari, E. N., & Wulandari, N. (2023). Ketepatan Obat, Dosis, dan Potensi Interaksi Obat pada Pasien Kanker Paru di Rumah Sakit X Jawa Barat Periode 2019-2021. *Jurnal Sains Dan Kesehatan*, *5*(3), 313–323. https://doi.org/10.25026/jsk.v5i3.1754

Sathyanarayanan, S., & Tantri, B. R. (2024). *Confusion Matrix-Based Performance Evaluation Metrics*. *27*(4).

Schapire, R. E. (1990). *The Strength of Weak Learnability*. *227*, 197–227.

Schapire, R. E. (2013). Explaining adaboost. *Empirical Inference: Festschrift in Honor of Vladimir N. Vapnik*, 37–52. https://doi.org/10.1007/978-3-642-41136-6_5

Shihab, M. Q. (2022). *Tafsir Al-Mishbah*.

Sinaga, R. B., Widiyanto, D., & Wahyono, B. T. (2022). Deteksi Dini Penyakit Kanker Paru dengan Gabungan Algoritma Adaboost dan Random Forest. *Seminar Nasional Mahasiswa Ilmu Komputer Dan Aplikasinya (SENAMIKA)*, 1–10. https://www.kaggle.com/datasets/mysarahmadbhat/lung-cancer

Sukmana, F., & Rozi, F. (2017). Rekomendasi Solusi Pada Sistem Computer Maintenance Management System Menggunakan Association Rule , Fisher Exact Test One Side P-Value Dan Double One Side P-Value. *4*(4), 213–220. https://doi.org/10.25126/jtiik.201744368

Syafika, V. A. N., & Karisma, R. D. L. N. (2023). Implementasi Support Vector Machine (SVM) dalam Penentuan Klasifikasi Indeks Khusus Penanganan Stunting di Indonesia. *Seminar Nasional Official Statistics*, *2023*(1), 267–276. https://doi.org/10.34123/semnasoffstat.v2023i1.1595

Wahdah, A. S., Windarti, I., Setyaningrum, E., dkk. (2024). *Tinjauan Pustaka : Risiko Kanker Paru pada Pasien dengan Riwayat Tuberculosis Paru dan Penyakit Paru Obstruktif Kronik (PPOK) Article Review : Risk of Lung Cancer in Patients with History of Pulmonary Tuberculosis and Chronic Obstructive Pulmonary Disea*. *14*(1), 1845–1850.

Wahyudin, Rismaningsih, F., Hernaeny, U., dkk. (2022). *Pengantar Statistika 2* (S. Haryanti (ed.)). Penerbit Media Sains Indonesia.

Wang, R. (2012). AdaBoost for Feature Selection , Classification and Its Relation with SVM * , A Review. *Physics Procedia*, *25*, 800–807. https://doi.org/10.1016/j.phpro.2012.03.160

Widyawati, A. S., Fitrianto, A., & Silvianti, P. (2025). *Performance of Multivariate Missing Data Imputation Methods on Climate Data*. *9*(6).

LAMPIRAN

Lampiran 1 Data Pasien Kanker Paru-Paru Tahun 2020-2025

No	Usia	Jenis Kelamin	Status Merokok	Komorbiditas	Riwayat Keluarga	Stadium
1	25	2	2	1	2	-1
2	42	2	2	1	2	-1
3	72	1	1	2	2	-1
4	57	1	1	1	2	-1
5	61	1	1	2	1	-1
6	78	1	1	2	2	-1
…	…	…	…	…	…	…
127	36	1	1	2	2	-1
128	49	1	1	2	2	-1
129	74	1	1	2	2	-1
130	33	2	1	2	2	-1
131	69	2	1	1	2	-1
132	47	2	1	2	2	-1

Lampiran 2 Standardisasi Data Usia

No	Usia	Standardisai
1	25	-3.11
2	42	-1.62
3	72	1.01
4	57	-0.31
5	61	0.04
6	78	1.53
…	…	…
127	36	-2.15
128	49	-1.01
129	74	1.18
130	33	-2.41
131	69	0.74
132	47	-1.18

Lampiran 3 *Confusion Matrix* Parameter 0.01

Confusion Matrix		Aktual	
		Stadium III	Stadium IV
Prediksi	Stadium III	0	0
	Stadium IV	2	11

Lampiran 4 *Confusion Matrix* Parameter 0.1

Confusion Matrix		Aktual	
		Stadium III	Stadium IV
Prediksi	Stadium III	0	0
	Stadium IV	2	11

Lampiran 5 *Confusion Matrix* Parameter 1

Confusion Matrix		Aktual	
		Stadium III	Stadium IV
Prediksi	Stadium III	1	3
	Stadium IV	1	8

Lampiran 6 *Confusion Matrix* Parameter 10

Confusion Matrix		Aktual	
		Stadium III	Stadium IV
Prediksi	Stadium III	1	2
	Stadium IV	1	9

Lampiran 7 *Confusion Matrix* Parameter 100

Confusion Matrix		Aktual	
		Stadium III	Stadium IV
Prediksi	Stadium III	1	2
	Stadium IV	1	9

Lampiran 8 *Syntax* Model AdaBoost-SVM dengan RStudio

```
# IMPORT LIBRARY
library(readxl)
library(writexl)
library(dplyr)
library(e1071)
library(caret)


# INPUT DATA
Data <- read_excel("C:/Users/USER/Documents/data
skripsi.xlsx")


# Mengecek struktur dan missing values
str(data)
colSums(is.na(data))

# Membentuk indikator missing data
data$R_merokok <- ifelse(is.na(data$`Status Merokok`), 1, 0)
data$R_komorbid <- ifelse(is.na(data$Komorbiditas), 1, 0)
data$R_riwayat <- ifelse(is.na(data$`Riwayat Keluarga`), 1,
0)
data$R_stadium <- ifelse(is.na(data$Stadium), 1, 0)

# Membuat tabel kontingensi
tab1 <- table(data$R_merokok, data$`Jenis Kelamin`)
tab2 <- table(data$R_komorbid, data$`Jenis Kelamin`)
tab3 <- table(data$R_riwayat, data$`Jenis Kelamin`)
tab4 <- table(data$R_stadium, data$`Jenis Kelamin`)

# Uji chi-square
uji1 <- chisq.test(tab1)
uji2 <- chisq.test(tab2)
uji3 <- chisq.test(tab3)
uji4 <- chisq.test(tab4)

# Melihat frekuensi harapan
uji1$expected
uji2$expected
uji3$expected
uji4$expected

# Uji hubungan missing data dengan variabel kategorik
chisq.test(tab1)
fisher.test(tab2)
chisq.test(tab3)
chisq.test(tab4)
```

```
# Uji hubungan missing data dengan variabel numerik
wilcox.test(usia ~ R_stadium, data = data)
wilcox.test(usia ~ R_merokok, data = data)
wilcox.test(usia ~ R_komorbid, data = data)
wilcox.test(usia ~ R_riwayat, data = data)


set.seed(123) # menjaga konsistensi hasil pengacakan data

# IMPUTASI MISSING DATA DENGAN MODUS
# Fungsi untuk imputasi dengan modus
impute_mode <- function(x) {
  mode_value <- as.numeric(names(sort(table(x), decreasing =
TRUE))[1])
  x[is.na(x)] <- mode_value
  return(x)
}

# Mengaplikasikan imputasi modus ke seluruh kolom
data_clean <- data %>%
  mutate(across(everything(), impute_mode))

head(data_clean) # Menampilkan data setelah imputasi

# STANDARDISASI DATA
data_clean$usia <- scale(data_clean$usia)

head(data_clean) # Menampilkan data setelah standardisasi

# MENGUBAH TIPE VARIABEL TARGET MENJADI FAKTOR
data_clean$Stadium <- factor(
  data_clean$Stadium,
  levels = c(-1, 1),
  labels = c("-1", "1")
)

# DATA TRAINING DAN TESTING
set.seed(9191)
n=round(nrow(data_clean)*0.90);n  # persentase data training
sampel = sample(1:nrow(data_clean),n)  # agar data random
set.seed(1919)
train_data = data_clean[sampel,]  # Membentuk data training
test_data = data_clean[-sampel,]  # Membentuk data testing

# MENYIMPAN DATA TRAINING DAN DATA TESTING
write_xlsx(train_data, "train_data4 (90).xlsx")
write_xlsx(test_data, "test_data4 (10).xlsx")

# MEMISAHKAN FITUR DAN VARIABEL TARGET
train_X <- train_data %>% select(-Stadium)
train_y <- train_data$Stadium
test_X <- test_data %>% select(-Stadium)
test_y <- test_data$Stadium
```

```
# MEMBENTUK MODEL ADABOOST-SVM
set.seed(123)

adaboost_svm_train <- function(train_data, T = 10) {

  X <- train_X
  y <- train_y
  m <- nrow(train_data) # jumlah observasi data training

 # mengubah label kelas menjadi numerik untuk AdaBoost
  y_num <- ifelse(y == "1", 1, -1)

  D <- rep(1/m, m) # inisialisasi bobot awal

  models <- list() # menyimpan model SVM tiap iterasi
  alpha  <- numeric() # menyimpan alpha tiap iterasi

  for (t in 1:T) {

    cat("\nIterasi", t, "\n")

   #bootstrap berbobot
    idx <- sample(1:m, m, replace = TRUE, prob = D)
    boot <- train_data[idx, ]

   # SVM kernel linear
    model <- svm(
      Stadium ~ .,
      data = boot,
      kernel = "linear",
      cost = 10,
      scale = FALSE
    )

    # prediksi
    full_pred <- predict(model, X)
    full_pred_num <- ifelse(full_pred == "1", 1, -1)

    # weighted error
    err_t <- sum(D * (full_pred != y))
    cat("Weighted Error (epsilon):", round(err_t, 4), "\n")

    if (err_t >= 0.5) break # Menghentikan iterasi jika
error ≥ 0.5

    err_t <- min(max(err_t, 1e-10), 0.499)  # Menjaga nilai
error agar tidak ekstrem

    # Menghitung bobot pengaruh model (alpha_t)
    alpha_t <- 0.5 * log((1 - err_t) / err_t)
    alpha <- c(alpha, alpha_t)

    cat("Alpha (alpha_t):", round(alpha_t, 4), "\n")
```

```
    cat("Bobot D sebelum update:\n")
    print(round(D, 4))

    # Memperbarui bobot berdasarkan hasil prediksi
    D <- D * exp(-alpha_t * y_num * full_pred_num)
    D <- D / sum(D)

    cat("Bobot D sesudah update:\n")
    print(round(D, 4))
    cat("Total bobot:", sum(D), "\n")

    models[[t]] <- model # Menyimpan model SVM hasil iterasi
ke-t
  }

  list(models = models, alpha = alpha)
}

# PREDIKSI AKHIR DARI GABUNGAN SELURUH MODEL
adaboost_svm_predict <- function(model, X, threshold = -0.5)
{

  agg <- rep(0, nrow(X))

  for (t in seq_along(model$models)) {
    p <- predict(model$models[[t]], X)
    p_num <- ifelse(p == "1", 1, -1)
    agg <- agg + model$alpha[t] * p_num
  }

  factor(ifelse(agg >= threshold, "1", "-1"), levels = c("-
1","1"))
}

# MEMBANGUN MODEL ADABOOST-SVM
model_final <- adaboost_svm_train(train_data, T = 10)

# MELAKUKAN PREDIKSI PADA DATA TESTING
pred_test <- adaboost_svm_predict(model_final, test_X,
threshold = -0.9)

table(pred_test)# menghitung jumlah hasil prediksi
pred_test # menampilkan hasil prediksi

# EVALUASI K-FOLD CROSS VALIDATION
set.seed(123)
k <- 5 # Menentukan jumlah fold
folds <- createFolds(train_data$Stadium, k = k)

cv_acc <- numeric(k)

for (i in 1:k) {
```

```
  test_idx <- folds[[i]]

  train_cv <- train_data[-test_idx, ]
  test_cv  <- train_data[test_idx, ]

  model_cv <- adaboost_svm_train(train_cv, T = 10)

  pred_cv <- adaboost_svm_predict(model_cv,
                                  test_cv %>% select(-
Stadium))

  pred_cv <- factor(pred_cv, levels=c("-1","1"))
  test_cv$Stadium <- factor(test_cv$Stadium, levels=c("-
1","1"))

  cm <- confusionMatrix(pred_cv, test_cv$Stadium,
positive="1")
  print(cm)

  cv_acc[i] <- cm$overall["Accuracy"]
}

mean(cv_acc) # Menghitung rata-rata akurasi dari seluruh fold
sd(cv_acc) # Menghitung standar deviasi

# MENGUBAH NILAI PREDIKSI MENJADI FAKTOR
pred_class <- factor(pred_test, levels=c("-1","1"))
test_y <- factor(test_y, levels=c("-1","1"))

# EVALUASI CONFUSION MATRIX
cm <- confusionMatrix(pred_class, test_y, positive="1")
cm

conf_matrix <- cm$table
# Mengambil tabel confusion matrix (TP, TN, FP, FN)
TP <- conf_matrix["1","1"]
TN <- conf_matrix["-1","-1"]
FP <- conf_matrix["1","-1"]
FN <- conf_matrix["-1","1"]

# Hitung Metrik Evaluasi
akurasi <- (TP+TN)/(TP+TN+FP+FN)
recall <- ifelse((TP+FN)==0,0,TP/(TP+FN))
presisi <- ifelse((TP+FP)==0,0,TP/(TP+FP))
f1_score <- ifelse((recall+presisi)==0,0,
                   2*(recall*presisi)/(recall+presisi))

# Menampilkan Hasil Evaluasi
cat("\n--- HASIL EVALUASI ADA BOOST + SVM ---\n")
cat("Akurasi :",round(akurasi,4),"\n")
cat("Recall :",round(recall,4),"\n")
cat("Presisi :",round(presisi,4),"\n")
cat("F1 :",round(f1_score,4),"\n")
```

RIWAYAT HIDUP

Aulia Nursafitri, lahir di Jakarta, 24 Desember 2003. Penulis merupakan putri sulung dari dua bersaudara dari pasangan Bapak Agus Irawan dan Ibu Sri Mulyani. Penulis telah menempuh pendidikan formal yang dimulai di RA Attaqwa 21 Bekasi dan tamat pada tahun 2010. Selanjutnya, penulis melanjutkan pendidikan dasar di SDIT Attaqwa Pusat Bekasi dan berhasil diselesaikan pada tahun 2016. Kemudian, penulis melanjutkan pendidikan menengah pertama di SMPIT Attaqwa Pusat Bekasi dan lulus pada tahun 2019. Setelah itu, penulis melanjutkan pendidikan menengah atas di MAN 1 Kota Bekasi dan lulus pada tahun 2022. Pada tahun yang sama, penulis melanjutkan pendidikan perguruan tinggi negeri di Universitas Islam Negeri Maulana Malik Ibrahim Malang, Fakultas Sains dan Teknologi, Program Studi Matematika. Selama menjalani pendidikan di perguruan tinggi, penulis aktif mengikuti kegiatan organisasi kemahasiswaan, yaitu sebagai anggota HMPS ‘Integral’ Matematika Divisi *External Public Relation* pada periode tahun 2023. Kegiatan tersebut memberikan penulis banyak pengalaman dalam mengembangkan kemampuan komunikasi, kerja sama dalam tim, serta rasa tanggung jawab dalam berorganisasi. Selain itu, penulis juga aktif terlibat dalam berbagai kegiatan kepanitiaan kampus yang memberikan pengalaman untuk mengembangkan kemampuan pengelolaan administrasi sebuah acara. Berbagai pengalaman tersebut dapat menjadi sarana pembelajaran bagi penulis dalam mengembangkan kemampuan akademik, profesional, serta keterampilan diri selama masa perkuliahan.

KEMENTERIAN AGAMA RI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG FAKULTAS SAINS DAN TEKNOLOGI
Jl. Gajayana No. 50 Malang 65144 Telp. / Fax. (0341) 558933

BUKTI KONSULTASI SKRIPSI

Nama	:	Aulia Nursafitri
NIM	:	220601110087
Fakultas / Jurusan	:	Sains dan Teknologi / Matematika
Judul Skripsi	:	Penerapan *Adaptive Boosting-Support Vector Machine* Pada Klasifikasi Pasien Kanker Paru-Paru Berdasarkan Faktor Risiko.
Pembimbing I	:	Ria Dhea Layla Nur Karisma, M.Si.
Pembimbing II	:	Juhari, M.Si.

No	Tanggal	Hal	Tanda Tanggan
1.	16 Juni 2025	Konsultasi Topik dan Data	1.
2.	27 Agustus 2025	Konsultasi Bab I	2.
3.	10 September 2025	Konsultasi Bab I, II, III	3.
4.	17 September 2025	Konsultasi Bab I, II, III	4.
5.	24 September 2025	Konsultasi Bab I, II, III	5.
6.	1 Oktober 2025	Konsultasi Bab I, II, III	6.
7.	3 Oktober 2025	Konsultasi Kajian Agama Bab I dan II	7.
8.	8 Oktober 2025	Konsultasi Bab I, II, III	8.
9.	14 Oktober 2025	ACC Kajian Agama Bab I dan II	9.
10.	22 oktober 2025	Konsultasi Bab I, II, III	10.
11.	29 Oktober 2025	ACC Bab I, II, dan III	11.
12.	29 Oktober 2025	ACC Seminar Proposal	12.
13.	12 November 2025	Konsultasi Revisi Seminar Proposal	13.

KEMENTERIAN AGAMA RI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG FAKULTAS SAINS DAN TEKNOLOGI
Jl. Gajayana No. 50 Malang 65144 Telp. / Fax. (0341) 558933

14.	26 November 2025	Konsultasi Revisi Seminar Proposal	14.
15.	3 Februari 2026	Konsultasi Bab IV dan V	15.
16.	13 Februari 2026	Konsultasi Bab IV dan V	16.
17.	24 Februari 2026	Konsultasi Bab IV dan V	17.
18.	27 Februari 2026	Konsultasi Kajian Agama Bab IV	18.
19.	27 Februari 2026	ACC Kajian Agama Bab IV	19.
20.	2 Maret 2026	ACC Bab IV dan V	20.
21.	2 Maret 2026	ACC Seminar Hasil	21.
22.	14 April 2026	Konsultasi Revisi Seminar Hasil	22.
23.	28 April 2026	ACC Sidang Skripsi	23.
24.	11 Mei 2026	ACC Keseluruhan	24.

Malang, 18 Mei 2026
Mengetahui,
Ketua Program Studi Matematika

Dr. Fachrur Rozi, M.Si
NIP. 19800527 200801 1 012