KLASIFIKASI KANKER SERVIKS MENGGUNAKAN METODE *SUPPORT VECTOR MACHINE* BERBASIS *PRINCIPAL COMPONENT ANALYSIS*

SKRIPSI

Oleh:
NI'MATUL WAFIROH
NIM. 200605110006

PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2024

KLASIFIKASI KANKER SERVIKS MENGGUNAKAN METODE *SUPPORT VECTOR MACHINE* BERBASIS *PRINCIPAL COMPONENT ANALYSIS*

SKRIPSI

Diajukan Kepada:
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang
Untuk memenuhi Salah Satu Persyaratan Dalam
Memperoleh Gelar Sarjana Komputer (S.Kom)

Oleh:
NI'MATUL WAFIROH
NIM. 200605110006

PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM MALANG
2024

HALAMAN PERSETUJUAN

KLASIFIKASI KANKER SERVIKS MENGGUNAKAN METODE *SUPPORT VECTOR MACHINE* BERBASIS *PRINCIPAL COMPONENT ANALYSIS*

SKRIPSI

Oleh:
NI'MATUL WAFIROH
NIM. 200605110006

Telah Diperiksa dan Disetujui untuk Diuji
Tanggal: 29 Mei 2024

Pembimbing I	Pembimbing II
Prof. Dr. Suhartono, M.Kom	Roro Inda Melani, M.T, M.Sc
NIP. 19680519 200312 1 001	NIP. 19780925 200501 2 008

Mengetahui,
Ketua Program Studi Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang

Dr. Fachrul Kurniawan, M.MT, IPM
NIP. 19771020 200912 1 001

HALAMAN PENGESAHAN

KLASIFIKASI KANKER SERVIKS MENGGUNAKAN METODE *SUPPORT VECTOR MACHINE* BERBASIS *PRINCIPAL COMPONENT ANALYSIS*

SKRIPSI

Oleh:
NI'MATUL WAFIROH
NIM. 200605110006

Telah Dipertahankan di Depan Dewan Penguji Skripsi
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan
Untuk Memperoleh Gelar Sarjana Komputer (S.Kom)
Tanggal: 6 Juni 2024

Susunan Dewan Penguji

Ketua Penguji	: Dr. Irwan Budi Santoso, M.Kom NIP. 19770103 201101 1 004	( )
Anggota Penguji I	: Agung Teguh Wibowo Almais, M.T NIP. 19860301 2023211 016	( )
Anggota Penguji II	: Prof. Dr. Suhartono, M.Kom NIP. 19680519 200312 1 001	( )
Anggota Penguji III	: Roro Inda Melani, M.T, M.Sc NIP. 19780925 200501 2 008	( )

Mengetahui dan Mengesahkan,
Ketua Program Studi Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang

Fachrul Kurniawan, M.MT, IPM
NIP. 19771020 200912 1 001

PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini:

Nama	: Ni'matul Wafiroh
NIM	: 200605110006
Fakultas / Program Studi	: Sains dan Teknologi / Teknik Informatika
Judul Skripsi	: Klasifikasi Kanker Serviks Menggunakan Metode *Support Vector Machine* Berbasis *Principal Component Analysis*

Menyatakan dengan sebenarnya bahwa Skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri, bukan merupakan pengambil alihan data, tulisan, atau pikiran orang lain yang saya akui sebagai hasil tulisan atau pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka.

Apabila dikemudian hari terbukti atau dapat dibuktikan skripsi ini merupakan hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

MOTTO

*“*Lelah boleh, nyerah jangan 😊 *”*

“Tiada Kelezatan kecuali habis berpayah-payah”

~Mahfudlot~

HALAMAN PERSEMBAHAN

Skripsi ini dipersembahkan untuk kedua orang tua saya yaitu Bapak Abdul Rohman dan Ibu Sumiati sebagai bentuk tanggung jawab saya kepada orang tua saya. Beliau berdua yang senantiasa berjuang agar anaknya bisa melanjutkan dibangku kuliah, semoga beliau berdua diberikan balasan yang setimpal atas perjuangannya. Dan tak lupa untuk kakakku tersayang Mbak Halimatus Sya'diyah yang selalu ada buat adik tercintanya ini.

KATA PENGANTAR

Assalamualaikum Warahmatullahi Waabarakatuh

Dengan menyebut nama Allah yang Maha Pengasih lagi Maha Penyayang, segala puji bagi Allah Tuhan Semesta Alam. Atas segala rahmat dan karunia-Nya sehingga penulis mampu menyelesaikan skripsi dengan judul "Klasifikasi Kanker Serviks Menggunakan Metode *Support Vector Machine* Berbasis *Principal Component Analysis*". Sholawat serta salam semoga selalu tercurah limpkahkan kepada Nabi Muhammad SAW yang telah membimbing umatnya dari zaman *jahiliyah* menuju zaman *islamiyah* yakni *addinul islam*.

Dalam penyusunan skripsi ini, saya merasakan karunia dan bimbingan Allah SWT dengan segala keberhasilan dan ekmajuan yang telah saya raih tidak lepas dari rahmat-Nya serta petunjuk-Nya yang tak tergantikan. Penyusunan skripsi ini juga tidak lepas dari peran dan dukungan banyak pihak yang tak ternilai. Oleh karena itu, penulis ingin mengucapkan terimakasih kepada:

1. Prof. Dr. H. M. Zainuddin, M.A., selaku rektor Universitas Islam Negeri Maulana Malik Ibrahim Malang.
2. Prof. Dr. Sri Harini, M.Si., selaku dekan Fakultas Sains dan Teknologi.
3. Dr. Fachrul Kurniawan, M.MT selaku Ketua Program Studi Teknik Informatika.
4. Prof Suhartono, M.Kom dan Roro Inda Melani, M.T, M.Sc selaku dosen pembimbing yang dengan penuh kesabaran dan kebijaksanaan telah membimbing saya dalam menyelesaikan skripsi ini. Bimbingan, nasihat, dan

pengarahan yang diberikan sangat berarti bagi kemajuan dan kelancaran penelitian ini.

5. Dr. Irwan Budi Santoso, M.Kom dan Agung Teguh Wibowo Almais, M.T selaku dosen penguji yang dengan sabar telah memberi arahan dan saran dalam proses penelitian ini.
6. Segenap civitas akademika Program Studi Teknik Informatika, terutama seluruh dosen. Terimakasih atas ilmu dan bimbingan yang telah diberikan selama masa perkuliahan ini.
7. Bapak Abdul Rohman selaku bapak dari penulis yang tak kenal pagi, siang, dan malam hanya untuk bekerja demi anaknya bisa hidup di perantauan untuk menyelesaikan studinya di kampus impian anaknya. Serta Ibu Sumiati yang tak henti-henti memberi nasihat, masukan, serta doa yang selalu diberikan kepada penulis ini.
8. Teman-teman dekat penulis yang menjadi saksi skripsi ini dibuat dan teman-teman Program Studi Teknik Informatika Angkatan 2020 “Integer” yang sama-sama mengejar gelar S.Kom.
9. Teman-teman Lembaga Tinggi Pesantren Luhur Malang yang selalu menemani penulis dari bangun tidur sampai tidur lagi dan selalu memberi semangat agar bisa menyelesaikan skripsi ini.
10. Semua pihak yang telah membantu penulis dalam menyelesaikan skripsi ini, baik secara langsung maupun tidak langsung yang tidak dapat disebutkan satu per satu.
11. Penulis sendiri yang dapat menyelesaikan skripsi dan tanggung jawab penulis.

Akhir kata, penulis menyadari bahwa penulisan pada skripsi ini masih jauh dari kata sempurna. Saya berdoa semoga skripsi ini dapat diterima sebagai amal ibadah yang ikhlas dan bermanfaat disisi Allah SWT. Semoga karya ini menjadi bentuk kontribusi dalam rangka memperkokoh dan mengembangkan ilmu pengetahuan dan menjalankan tugas sebagai hamba Allah SWT yang bertanggung jawab.

Malang, 12 Juni 2024

Penulis

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

ABSTRAK

Wafiroh, Ni'matul. 2024. ***Klasifikasi Kanker Serviks Menggunakan Metode Support Vector Machine Berbasis Principal Component Analysis***. Skripsi. Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) Prof Suhartono, M.Kom (II) Roro Inda Melani, M.T, M.Sc.

.

Kata kunci: Klasifikasi Kanker Serviks, *Support Vector Machine, Principal Analysis Component*.

Kanker serviks merupakan salah satu penyakit yang paling umum dan mematikan bagi perempuan di seluruh dunia, terutama di negara-negara berkembang. Untuk mengatasi masalah ini, pengembangan model klasifikasi berdasarkan data demografis, kebiasaan, dan riwayat medis menjadi penting. Penelitian ini menggunakan dataset *Cervical cancer (Risk Factors)* dari UCI *Machine Learning*, yang terdiri dari 32 faktor risiko dan 4 target kanker serviks. Fokus penelitian adalah pada klasifikasi kanker serviks menggunakan metode *Support Vector Machine* (SVM) berbasis *Principal Component Analysis* (PCA). Penelitian ini juga menerapkan teknik *Synthetic Minority Over-sampling Technique* (SMOTE) untuk menangani ketidakseimbangan kelas dalam dataset. SMOTE digunakan untuk memperbanyak sampel kelas minoritas secara sintetis sehingga setiap kelas memiliki jumlah data yang seimbang. Setelah penyeimbangan kelas dengan SMOTE, PCA diterapkan untuk mereduksi dimensi dataset dengan menyederhanakan dan menghilangkan fitur yang tidak relevan tanpa mengurangi informasi penting. Hasil penelitian menunjukkan bahwa kombinasi SMOTE dan PCA dengan SVM menghasilkan model klasifikasi yang lebih akurat dan stabil. Penggunaan PCA setelah SMOTE membantu dalam mengurangi overfitting, yang terlihat ketika PCA digunakan tanpa SMOTE. Kombinasi terbaik ditemukan dengan menggunakan SMOTE untuk penyeimbangan kelas dan SVM untuk klasifikasi, yang secara signifikan meningkatkan akurasi dan kemampuan generalisasi model. Penelitian ini berkontribusi pada pengembangan metode klasifikasi yang lebih efektif dan efisien untuk diagnosis kanker serviks, menekankan pentingnya penyeimbangan kelas dan pengurangan dimensi dalam meningkatkan kinerja model pembelajaran mesin.

ABSTRACT

Wafiroh, Ni'matul. 2024. Classification of Cervical Cancer Using the Support Vector Machine Method Based on Principal Component Analysis. Thesis. Department of Informatics Engineering, Faculty of Science and Technology, Maulana Malik Ibrahim State Islamic University, Malang. Supervisor: (I) Prof Suhartono, M.Kom (II) Roro Inda Melani, M.T, M.Sc.

Cervical cancer is one of the most common and deadly diseases for women throughout the world, especially in developing countries. To overcome this problem, the development of classification models based on demographic data, habits, and medical history is important. This research uses the Cervical Cancer (Risk Factors) dataset from UCI Machine Learning, which consists of 32 risk factors and 4 targets for cervical cancer. The focus of the research is on cervical cancer classification using the Support Vector Machine (SVM) method based on Principal Component Analysis (PCA). This research also applies the Synthetic Minority Over-sampling Technique (SMOTE) technique to handle class continuity in the dataset. SMOTE is used to synthetically increase minority class samples so that each class has a balanced amount of data. After class balancing with SMOTE, PCA is applied to reduce the dimensionality of the dataset with the help of and remove irrelevant features without reducing important information. The research results show that the combination of SMOTE and PCA with SVM produces a more accurate and stable classification model. The use of PCA after SMOTE helps in reducing overfitting, which is seen when PCA is used without SMOTE. The best combination was found to use SMOTE for class balancing and SVM for classification, which significantly improved the accuracy and generalization ability of the model. This research contributes to the development of more effective and efficient classification methods for cervical cancer diagnosis, the importance of class balancing and dimensionality reduction in improving the performance of machine learning models.

Keywords: Classification Cervical Cancer, *Support Vector Machine, Principal Analysis Component*.

وفيره، نعمتول. 2024 .تصنيف سرطان عنق الرحم باستخدام طريقة آلة ناقل الدعم بناءً على تحليل المكونات الرئيسية .أُطرُوحَة قسم :الهندس المعلوماتية، كلية العلوم والتكنولوجيا، جامعة مولانا مالك إبراهيم الإسلامية الحكومية، مالانج المشرف الاول الاستاذ الدكتورسوهارتونوالمجستيركومفوتير. المشرف الثاني: رورو اندى ميلاني المجستير تيكنولوغي، المجستير في العلوم.

الكلمات المفتاحية: تصنيف سرطان عنق الرحم، جهاز ناقل الدعم، مكون التحليل الرئيسي

يعد سرطان عنق الرحم أحد أكثر الأمراض شيوعًا وفتكًا بالنساء في جميع أنحاء العالم، وخاصة في البلدان النامية. للتغلب على هذه المشكلة، من المهم تطوير نماذج التصنيف بناءً على البيانات الديموغرافية والعادات والتاريخ الطبي. يستخدم هذا البحث مجموعة بيانات سرطان عنق الرحم عوامل الخطر من **UCI Machine Learning**، والتي تتكون من **32** عامل خطر و**4** أهداف لسرطان عنق الرحم. ينصب تركيز البحث على تصنيف سرطان عنق الرحم باستخدام طريقة آلة ناقل الدعم **SVM** القائمة على تحليل المكونات الرئيسية **PCA**. يطبق هذا البحث أيضًا تقنية تقنية أخذ العينات الزائدة للأقليات الاصطناعية **SMOTE** للتعامل مع استمرارية الفصل في مجموعة البيانات. يتم استخدام **SMOTE** لزيادة عينات فئات الأقليات بشكل صناعي بحيث يكون لكل فئة كمية متوازنة من البيانات. بعد موازنة الفئة مع **SMOTE**، يتم تطبيق **PCA** لتقليل أبعاد مجموعة البيانات بمساعدة وإزالة الميزات غير ذات الصلة دون تقليل المعلومات المهمة. تظهر نتائج البحث أن الجمع بين **SMOTE** و**PCA** مع **SVM** ينتج تصنيفًا أكثر دقة واستقرارًا للنماذج. يساعد استخدام PCA بعد SMOTE في تقليل التجهيز الزائد، وهو ما يظهر عند استخدام **PCA** بدون **SMOTE**. تم العثور على أفضل مزيج لاستخدام **SMOTE** لموازنة الفئة و**SVM** للتصنيف، مما أدى إلى تحسين دقة النموذج وقدرته على التعميم بشكل كبير. يساهم هذا البحث في تطوير طرق تصنيف أكثر فعالية وكفاءة لتشخيص سرطان عنق الرحم، وأهمية التوازن الطبقي وتقليل الأبعاد في تحسين أداء نماذج التعلم الآلي.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Menurut WHO (*World Health Organization*) kanker adalah sekelompok besar penyakit yang dapat dimulai dari semua organ atau jaringan tubuh ketika sel-sel abnormal tumbuh secara tidak terkendali, melampaui batas, dan biasanya menyerang bagian tubuh yang berdekatan dan/atau menyebar ke organ lain. Kondisi ini biasanya disebabkan oleh kerusakan dan pertumbuhan yang tidak normal pada gen yang mengatur regenerasi sel (N. P. W. P. Sari & Bura Mare, 2022).

Salah satu kanker yang mempengaruhi kesehatan wanita adalah kanker serviks. Kanker serviks merupakan kanker yang menyerang organ intim wanita yang terletak di leher rahim yang berada pada bagian paling bawah pada uterus. Kanker serviks menjadi penyebab kanker terbesar nomor dua bagi wanita di seluruh dunia. Data kanker global mengungkapkan bahwa kanker serviks adalah penyakit paling umum ke empat yang menyerang perempuan dengan tingkat kematian sekitar 90% di negara-negara terbelakang dan berkembang karena minimnya atau tidak ada pengetahuan publik tentang penyebab dan dampak yang ditimbulkan dari kanker serviks ini (Andrian *et al*., 2020). Penyebab utamanya adalah HPV (*Human Papilloma Virus*), yang merupakan salah satu virus yang ditularkan melalui seksual. HPV adalah virus yang menginfeksi kulit (epidermis) dan membran mukosa manusia, seperti mukosa oral, esofagus, laring, trakea, konjungtiva, genital, dan anus (Field & Gilson, 2018). Virus ini dapat membahayakan sel-sel serviks, sel

skuamosa, dan sel kelenjar. Faktor-faktor yang berhubungan dengan perkembangan kanker serviks termasuk aktivitas seksual yang dimulai dari usia muda (kurang dari 16 tahun), jumlah total pasangan seksual yang tinggi (lebih dari empat), dan riwayat kutil kelamin (Andrian *et al*., 2020). Sebenarnya vaksin untuk virus seperti HPV ini sudah tersedia di pasaran, namun kurangnya kesadaran, pengetahuan, dan juga sosialisasi yang baik, maka pencegahan ini tidak umum digunakan. Hal paling penting yang harus dilakukan untuk mencegah terjadinya kanker serviks yaitu melakukan deteksi dini dan perawatan yang tepat untuk mengurangi insiden dan resiko kematian yang bertujuan agar pasien yang menderita penyakit kanker serviks dapat mencegah dengan obat-obatan dan perawatan khusus sehingga dapat sembuh. Seperti firman Allah SWT dalam QS. *Al-Isra'* ayat 82:

وَنُنَزِّلُ مِنَ الْقُرْاٰنِ مَا هُوَ شِفَاۤءٌ وَّرَحْمَةٌ لِّلْمُؤْمِنِيْنَۙ وَلَا يَزِيْدُ الظّٰلِمِيْنَ اِلَّا خَسَارًا

"Dan Kami turunkan dari Al-Qur'an (sesuatu) yang menjadi penawar dan rahmat bagi orang yang beriman, sedangkan bagi orang yang zalim (Al-Qur'an itu) hanya akan menambah kerugian."(Q.S. Al-Isra': 82)

Dari ayat tersebut dapat diketahui bahwa, Allah tidak semata-mata meringankan penyakit tetapi juga memberikan penawarnya. Selain itu fungsi dari Al-Quran adalah *As Syifa* yang berarti obat. Allah telah menganugerahkan Al-Quran sebagai penawar segala penyakit kepada hamba-hamba-Nya yang beriman. Saat sakit, sebagai hamba yang setia harus selalu berdoa dan mengupayakan kesembuhan. Seperti halnya kanker serviks, kanker serviks juga dapat disembuhkan jika pasien mendapat perawatan dan pengobatan khusus yang tepat waktu. Oleh

karena itu, diperlukan pemeriksaan rutin untuk mendeteksi kanker secara dini pada daerah serviks.

Terdapat beberapa macam pemeriksaan laboratorium yang dapat dilakukan untuk diagnosis kanker serviks antara lain: 1) Pemeriksaan pap smear (*Cytology*), yaitu pemeriksaan dengan pengambilan lapisan dai permukaan leher rahim untuk menilai perubahan bentuk sel; 2) Pemeriksaan *Schiller* atau sering dikenal dengan IVA (Inspeksi Visual dengan Asam Asetat), yaitu pemeriksaan dengan menggunakan larutan iodium untuk mengetahui perubahan warna jaringan yang mengalami kelainan; 3) Pemeriksaan kolposkopi atau disebut dengan *Hinselmann*, yaitu pemeriksaan dengan menggunakan alat untuk menentukan terdapat daerah yang abnormal dan letak kelainannya (Kusumawati *et al*., 2016). Pada pap smear akan dilakukan pengambilan cairan sel leher rahim untuk dijadikan sampel pemeriksaan. Setelah itu jika terdapat perubahan sifat pada sel leher rahim, maka akan dilakukan pemeriksaan IVA dengan meneteskan cairan asam asetat pada bagian permukaan leher rahim. Jika ditemukan lesi kanker pada rahimnya maka akan menjalani pemeriksaan kolposkopi atau Hinselmann. Dan apabila masih ditemukan perubahan sel yang mencurigakan, maka dilanjutkan ke pemeriksaan selanjutnya yaitu *Biopsy*, untuk mendapatkan informasi lanjut tentang sifat dan tingkat keparahan perubahan sel tersebut.

Maka dari itu dengan berkembangnya teknologi, akan dibangun model klasifikasi berdasarkan data informasi demografis, kebiasaan, dan riwayat medis. Data tersebut diambil dari *UCI Machine Learning* dengan dataset *Cervical cancer (Risk Factors)*. Dataset *Cervical cancer (Risk Factors)* mempunyai parameter

target yang berbentuk kategori. Sehingga data ini cocok untuk model klasifikasi. Yang terdiri dari 36 label yaitu 32 faktor resiko dan 4 target, yaitu *Hinselmann, Schiller, Cytology,* dan *Biopsy*.

Dalam upaya untuk mengatasi masalah kanker serviks, teknologi berkembang seperti *machine learning* juga dapat berperan penting. *Machine learning* adalah teknik otomatis yang memungkinkan mesin memeriksa data dalam jumlah besar, menemukan pola, dan belajar dari data untuk membantu prediksi dan pengambilan keputusan (Wang & Siau, 2019). *Machine learning* dapat digunakan dalam analisis data medis untuk meningkatkan deteksi dini dan diagnosis penyakit tersebut. Terdapat beberapa macam algoritma machine learning seperti Regresi Linear, *Decision Tree, Random Forest, K-Nearest Neighbors* (KNN), *Naive Bayes*, *Support Vector Machine* (SVM*), Neural Networks*, dan algoritma *Clustering* seperti K-Means dan *Hierarchical Clustering* (Chollet, 2017). *Support Vector Machine* sering disebut dengan SVM. Pada penelitian ini peneliti akan menggunakan singkatan tersebut. Metode klasifikasi yang digunakan yaitu menggunakan metode SVM, karena metode ini dapat mengatasi masalah klasifikasi dan regresi dengan linear dan nonlinear sehingga dapat menjadi suatu kemampuan algoritma pembelajaran pada klasifikasi ataupun regresi (Puspitasari *et al*., 2018).

Dengan bantuan *machine learning*, sistem cerdas ini dapat dikembangkan untuk memberikan dukungan kepada profesional medis dalam mendiagnosis kanker serviks. Hal ini dapat memungkinkan deteksi dini yang lebih akurat dan efisien, sehingga dapat meningkatkan peluang kesembuhan dan mengurangi angka kematian yang disebabkan penyakit tersebut.

Pada penelitian ini akan membahas tentang Klasifikasi Kanker Serviks Menggunakan Metode *Support Vector Machine* berbasis *Principal Component Analysis*. Penelitian ini bertujuan untuk mengetahui performa dari metode *Support Vector Machine* berbasis *Principal Analysis Component* dalam melakukan klasifikasi tersebut. Performa yang baik dalam klasifikasi kanker serviks dapat berdampak langsung pada deteksi dini penyakit tersebut dan pengambilan tindakan medis yang tepat.

1.2 Pernyataan Masalah

Seberapa besar nilai performa dalam mengklasifikasikan kanker serviks dengan metode *Support Vector Machine* berbasis *Principal Component Analysis* berdasarkan data *Cervical cancer (Risk Factors)*?

1.3 Batasan Masalah

1. Data yang digunakan adalah data sekunder yang dapat diakses oleh *public* melalui *UCI Machine Learning Repository*: *Cervical cancer* (*Risk Factors*).
2. Menggunakan Metode *Support Vector Machine* dengan satu jenis kernel linear.
3. Mereduksi parameter dengan menggunakan *Principal Component Analysis* (PCA).

1.4 Tujuan Penelitian

Mengetahui nilai performa dalam mengklasifikasikan kanker serviks menggunakan metode *Support Vector Machine* berbasis *Principal Component Analysis* berdasarkan data *Cervical cancer (Risk Factors).*

1.5 Manfaat Penelitian

1. Memberikan wawasan kepada pembaca tentang penggunaan metode *Support Vector Machine* (SVM) dalam klasifikasi kanker serviks.
2. Diharapkan pada hasil klasifikasi dapat membantu mengidentifikasi pasien kanker serviks.

BAB II

STUDI PUSTAKA

2.1 Penelitian Terkait

Penelitian yang dilakukan oleh Gokulnath & Shantharajah (2019) dengan judul "*An optimized feature selection based on genetic approach and support vector machine for heart disease*" yang melakukan pemilihan fitur untuk mengoptimalkan metod*e Support Vector Machine* pada penyakit jantung. Pada penelitian menggunakan fitur GA-SVM dengan mengidentifikasi 7 fitur untuk mendapatkan hasil penyakit jantung. Hasil keakuratan dari SVM yaitu 88,34% dengan fitur yang sudah dipilih dan analisis ROC membuktikan kinerja pengklasifikasi SVM yang baik (Gokulnath & Shantharajah, 2019).

Penelitian yang dilakukan oleh Harimoorthy & Thangavelu (2021) dengan judul "*Multi-disease prediction model using improved SVM-radial bias technique in healthcare monitoring system*" yang memprediksi penyakit dengan mengurangi fitur kumpulan pada dataset penyakit ginjal kronis, diabetes dan penyakit jantung dengan menggunakan metode kernel bias SVM-Radial. Pada hasil akurasinya mendapatkan akurasi dari SVM-linear 96,7%, SVM-polynomial 96,7%, *Random Forest* 97,8%, *Decision Tree* 66,3%, dan SVM-radial 98,8% (Harimoorthy & Thangavelu, 2021).

Penelitian yang berjudul "Analisa Metode *Random Forest Tree* dan *KNearest Neighbor* dalam Mendeteksi Kanker Serviks" yang dilakukan oleh Andrian et al., (2020) dengan menggunakan dataset dari rumah sakit Hospital

Universitario de 7 Caracas di Caracas, Venezuela. Pembagian yang dilakukan dengan rasio 7.5 data training dan 2.5 data testing. Total entri dari dataset yang dipakai sebanyak 858 data. Hasil akurasi terbaik pada penelitian ini didapatkan dengan metode KNN yang sukses dengan tingkat akurasi 90.6% dan Random Forest mendapatkan tingkat akurasi 88.7%. Sistem prediksi ini yang digunakan untuk memprediksi hasil tes *screening* untuk pasien yang terkena kanker serviks berhasil dibangun dan dapat digunakan tenaga medis untuk memberi keputusan klinis terhadap pasiennya (Andrian et al., 2020).

Penilitian yang dilakukan oleh Zahras & Rustam (2018) dengan judul "*Cervical Cancer Risk Classification Based on Deep Convolutional Neural Network*". Data ini didapatkan dari dataset yang dikumpulkan di Rumah Sakit Universitario de Caracas di Caracas, Venezuela. Kumpulan data tersebut terdiri dari informasi demografis, kebiasaan, dan riwayat medis. Dataset ini memiliki 32 risiko dan empat variabel target yaitu *Hinselmann, Schiller, Cytology,* dan *Biopsy*. Keakuratan yang didapat dari keempat variabel target itu mendapatkan sararan yang lebih dari 90%. Cara ini dapat membantu tenaga medis dalam mengklasifikasi penyakit kanker serviks dengan mudah (Zahras & Rustam, 2018).

Penelitian yang dilakukan oleh Prasetyo et al (2023), dengan judul "*Comparative Study of Heart Failure Prediction Algorithm: Logistic Regression and SVM*". Data pada penelitian ini diperoleh dari *UCI Machine Learning Repository*. Pada penelitian ini membandingkan dua metode yaitu kinerja dari *Logistic Regression* dan *Support Vector Machine* untuk memprediksi gagal jantung. Hasil penelitian terbaik didapatkan oleh metode *Support Vector Machine* dengan

hasil akurasi 93,58% dan hasil dari metode *Logistic Regression* mendapatkan hasil akurasi 92, 95%. Nilai *Support Vector Machine* lebih tinggi 0,63% dari *Logistic Regression* (Prasetyo *et al*., 2023)

Penelitian yang dilakukan Anggoro & Novitaningrum (2021) dengan judul "*Comparison of Accuracy Level of Support Vector Machine (SVM) and K-Nearest Neighbors (KNN) Algorithms in Predicting Heart Disease*". Data penyakit jantung ini dibuat oleh Andras Janosi, MD dari Hungaria Institute of Cardiology Budapest, William Steinbrunn, MD dari University Hospital, Zurich, Swiss, Matthias Pfisterer, MD dari University Hospital, Basel, Swiss, dan Robert Detrano, MD, Ph.D. dari VA Medical Center, Long Beach Foundation dan Klinik Cleveland. Penelitian ini menggunakan perbandingan metode *Support Vector Machine* (SVM) dan *K-Nearest Neighbor* (KNN). Hasil pengujian menunjukkan bahwa pengujian algoritma SVM dengan normalisasi mempunyai hasil akurasi yang lebih baik dibandingkan dengan algoritma KNN baik dengan atau tanpa normalisasi tetap menghasilkan akurasi yang kurang baik. Hasil klasifikasi SVM tanpa normalisasi sebesar 84,61% dan dengan normalisasi sebesar 90,10%, sedangkan algoritma KNN menunjukkan akurasi sebesar 64,83% dan dengan normalisasi sebesar 81,31% (Anggoro & Novitaningrum, 2021).

Penelitian terkait diatas akan diringkas dalam bentuk tabel agar mudah dimengerti sebagai berikut:

Tabel 2. 1 Penelitian Terkait

No	Judul	Referensi	Metode Penelitian	Hasil Penelitian	Perbedaan
1.	"*An optimized feature selection based*	Gokulnath & Shantharajah	*Support Vector*	Hasil keakuratan dari SVM yaitu 88,34% dengan	Objek yang digunakan berbeda yaitu

No	Judul	Referensi	Metode Penelitian	Hasil Penelitian	Perbedaan
	on genetic approach and support vector machine for heart disease"		*Machine* (SVM)	fitur yang sudah dipilih dan analisis ROC membuktikan kinerja pengklasifikasi SVM yang baik	penyakit jantung. Pada penelitian ini menggunakan metode yang sama yaitu SVM dengan objek kanker serviks.
2.	"*Multi-disease prediction model using improved SVM-radial bias technique in healthcare monitoring system*"	Harimoorthy & Thangavelu	*SVM, Random Forest*, dan *Decision Tree*	Hasil akurasinya mendapatkan akurasi dari SVM-linear 96,7%, SVM-polynomial 96,7%, Random Forest 97,8%, Decision Tree 66,3%, dan SVM-radial 98,8%	Objek yang digunakan berbeda yaitu penyakit ginjal kronis, diabetes dan penyakit jantung. Pada penelitian ini menggunakan metode yang sama yaitu SVM dengan objek kanker serviks
3.	Analisa Metode *Random Forest Tree* dan *K-Nearest Neighbor* dalam Mendeteksi Kanker Serviks	Andrian *et al*., (2020)	*Random Forest Tree* dan *K-Nearest Neighbor*	Hasil akurasi terbaik pada penelitian ini didapatkan dengan metode KNN yang sukses dengan tingkat akurasi 90.6% dan Random Forest mendapatkan tingkat akurasi 88.7%	Metode yang digunakan *Random Forest Tree* dan *K-Nearest Neighbor*. Peneliti menggunakan metode *Support Vector Machine*
4.	*Cervical Cancer Risk Classification Based on Deep Convolutional Neural Network*	Zahras & Rustam (2018)	*Convolutional Neural Networks*	Keakuratan yang didapat dari keempat variabel target itu mendapatkan sararan yang lebih dari 90%. Cara ini dapat membantu tenaga medis dalam mengklasifikasi penyakit kanker serviks dengan mudah	Metode yang digunakan *Convolutional Neural Networks.* Peneliti menggunakan metode *Support Machine Learning*

No	Judul	Referensi	Metode Penelitian	Hasil Penelitian	Perbedaan
5.	"*Comparative Study of Heart Failure Prediction Algorithm: Logistic Regression and SVM*".	Prasetyo *et al*	*Logistic Regression and SVM*	Hasil penelitian terbaik didapatkan oleh metode Support Vector Machine dengan hasil akurasi 93,58% dan hasil dari metode Logistic Regression mendapatkan hasil akurasi 92, 95%. Nilai Support Vector Machine lebih tinggi 0,63% dari Logistic Regression	Objek yang digunakan berbeda yaitu penyakit hati. Pada penelitian ini menggunakan metode yang sama yaitu SVM dengan objek kanker serviks.
6.	"*Comparison of Accuracy Level of Support Vector Machine (SVM) and K-Nearest Neighbors (KNN) Algorithms in Predicting Heart Disease*"	Anggoro & Novitaningrum	*Support Vector Machine (SVM)* dan *K-Nearest Neighbors (KNN)*	Hasil klasifikasi SVM tanpa normalisasi sebesar 84,61% dan dengan normalisasi sebesar 90,10%, sedangkan algoritma KNN menunjukkan akurasi sebesar 64,83% dan dengan normalisasi sebesar 81,31%	Objek yang digunakan berbeda yaitu penyakit jantung. Pada penelitian ini menggunakan metode yang sama yaitu SVM dengan objek kanker serviks.

Dari penelitian terkait yang telah menggunakan beberapa metode yaitu *Support Vector Machine* (SVM), *K-Nearest Neighbors* (KNN), *Logistic Regression, Random Forest, dan Decision Tree.* Hasil menunjukkan bahwa metode SVM mendapatkan akurasi yang baik dalam pengklasifikasikan penyakit. Untuk itu peneliti akan menggunakan metode SVM dalam penelitian ini dengan objek kanker serviks. Hasil dari klasifikasi akan dievaluasi menggunakan *confusion matrix* untuk mengetahui performa dari sistem klasifikasi. Terdapat tiga pembaruan dalam

penelitian ini. Pertama, penggunaan dataset *Cervical cancer (Risk Factors)* yang open akses. Kedua, penggunaan metode SVM pada klasifikasi penyakit kanker serviks. Dan yang terakhir, menggunakan PCA untuk reduksi dimensinya.

2.2 Kanker

Penyakit kanker terjadi karena pertumbuhan sel yang tidak normal dan menyebar ke bagian tubuh lainnya, bahkan dapat menyebabkan kematian (Rahayuwati *et al*., 2020). Penyakit kanker ini penyakit yang mematikan di dunia, penyakit yang selalu bergerak dalam tubuh tanpa disadari manusia, dalam artian manusia tidak menyadari bahwa telah menderita kanker sampai kanker sudah masuk stadium lanjut. Faktor genetik, faktor karsinogenik (seperti bahan kimia, virus, radiasi, hormon, dan iritasi yang terus menerus), dan gaya hidup yang sering dilakukan yang dapat menyebabkan penyakit kanker (seperti merokok, pola makan tidak sehat, kurang aktivitas, dan alkohol) adalah beberapa faktor yang mempengaruhi penyakit kanker ini). Tetapi, gaya hidup dan makanan memiliki peluang yang lebih besar untuk menyebabkan kematian yang diakibatkan oleh penyakit kanker. Konsumsi buah dan sayuran yang tidak mencukupi kebutuhan tubuh, indeks massa tubuh yang tinggi, tidak aktif, merokok, dan penggunaan alkohol. Terutama merokok, yang merupakan penyebab kematian yang dapat dicegah yang mencakup hampir 70% di seluruh dunia (Rahayuwati *et al*., 2020).

2.3 Kanker Serviks

Kanker serviks adalah kanker yang terjadi pada sel leher rahim yang terletak pada organ reproduksi wanita yang merupakan pintu masuk ke arah rahim yang

terletak antara rahim dengan liang senggama. Bukti statistik menunjukkan bahwa kanker ini biasanya terjadi pada wanita yang berumur antara 20 sampai 30 tahun (Ellyzabeth, 2018). Menurut data WHO, pada tahun 2018 terdapat sekitar 570.000 kasus kanker serviks pada wanita di seluruh dunia dan sekitar 311.000 orang meninggal dikarenakan penyakit ini. Keganasan sel ini disebabkan oleh virus *Human Papiloma Virus* (HPV). HPV merupakan jenis virus DNA yang memiliki ukuran kecil dan tidak memiliki selubung yang seringkali menginfeksi pada bagian reproduksi (Evriarti & Yasmon, 2019). Munculnya virus HPV ini penularannya hampir seluruhnya diakibatkan oleh hubungan seksual (Dharma *et al*., 2020). Penyakit ini menjadi penyakit mematikan 11 bagi kaum wanita setelah kanker payudara dan kanker serviks telah menempati urutan kategori penyakit yang banyak menelan korban jiwa di dunia.

Di Indonesia, penyakit ini menempati urutan dua yang menelan korban jiwa bagi wanita karena lebih dari 70% kasus yang datang ke rumah sakit dengan keadaan Pada umumnya wanita menjadi sasaran empuk untuk terkena beberapa jenis kanker ganas, sebab pada dasarnya kaum wanita banyak yang tidak menjaga pola makan sehat, olah raga secara rutin, dan cara merawat tubuh dengan baik agar terhindar dari penyakit yang mematikan (Arifin *et al*., 2021). Beberapa gejala kanker serviks meliputi pendarahan ringan, pendarahan saat melakukan hubungan seksual, keputihan yang berlebihan, dan keputihan secara menopause (American Cancer Society, 2019). Perkembangan sel kanker pada leher rahim tidak terjadi secara cepat melainkan membutuhkan waktu, sekitar 10 tahun. Pada masa ini, wanita yang terkena kanker tidak menunjukkan gejala apapun dan akhirnya masuk

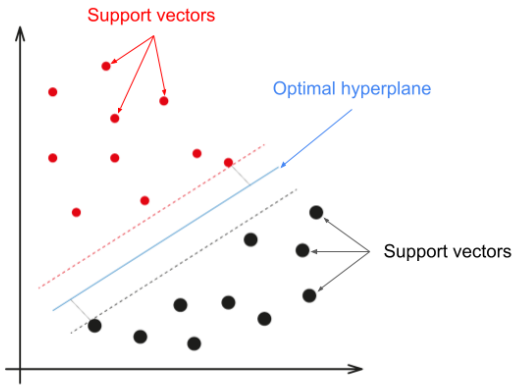
dalam stadium lanjut yang dapat menyebabkan kematian (Evriarti & Yasmon, 2019). Banyaknya kasus kematian itu juga dapat diakibatkan dari rendahnya cakupan deteksi dini kanker serviks diantaranya adalah tes *Papanicolaou* (PAP) *smear* dan Inspeksi Visual dengan Asam Asetat (IVA).

Dengan rendahnya cakupan deteksi dini, akan dilakukan klasifikasi kanker serviks dengan menggunakan empat cara tes. Tes pertama *Hinselmann*, merupakan suatu prosedur pemeriksaan pada wanita untuk mendeteksi perubahan sel pada leher rahim dengan menggunakan alat kolposkop. Tes kedua *Schiller*, merupakan suatu prosedur pemeriksaam pada wanita untuk mendeteksi perubahan sel pada leher rahim dengan melibatkan penggunaan larutan yodium. Tes ketiga *Cytology*, merupakan suatu prosedur pemeriksaam pada wanita untuk mendeteksi perubahan sel pada leher rahim dengan menggunakan alat spatula atau sikat. Dan tes terakhir yaitu *Biopsy*, merupakan suatu prosedur pemeriksaam pada wanita untuk mendeteksi perubahan sel pada leher rahim dengan menggunakan mikroskop.

2.4 *Support Vector Machine* (SVM)

Metode *Support Vector Machine* atau SVM dipublikasikan pada tahun 1992 oleh *Vapnik* bersama *Bernhard Boser, Isabelle Guyon*. Peningkatan akurasi klasifikasi dapat ditemukan dengan menggunakan SVM (Santoso *et al*., 2022). SVM adalah alat prediksi klasifikasi dan regresi yang berdasarkan teori *machine learning* untuk meningkatkan ketepatan pakurasi secara otomatis menghindari kecocokan data yang berlebihan (Pradana *et al*., 2022). SVM merupakan salah satu algoritma dalam *machine learning* yang berdasarkan prinsip *structural risk minimazation* yang bertujuan menciptakan *hyperlane* yang dapat memisahkan dua

kelas pada input space (Sutarto *et al*., 2019). Peningkatan akurasi klasifikasi dapat ditemukan dengan menggunakan SVM. *Hyperplane* adalah garis pembatas antara kedua kelas. Kelas yang dimaksud adalah output yang akan dihasilkan. Dengan bantuan *hyperplane*, untuk mengelompokkan data ke dalam kelas yang sesuai dengan karakteristiknya (Nugraha & Sabaruddin, 2021; Riadi, 2020).



Gambar 2. 1 Hyperplane

Dalam hal ini, dengan menggunakan ukuran baru, akan mendapatkan *hyperplane* optimal yang dapat digunakan untuk mengklasifikasikan data dari dua kelas secara *linear* atau dengan pemetaan nonlinear yang tepat ke dimensi yang lebih tinggi, dengan begitu data dari dua kelas dapat dipisah menggunakan *hyperplane* tersebut. SVM berupaya menemukan *hyperplane* yang paling baik, yaitu *hyperplane* yang terletak antara dua set objek yang masing-masing berasal dari kelas yang berbeda (Ritonga & Purwaningsih, 2018).

Untuk menemukan *hyperplane* yang optimal, SVM memerlukan solusi masalah optimasi kuadratik (QP, Quadratic *Programming*) yang kompleks.

Sequential Minimal Optimization (SMO) adalah algoritma yang digunakan untuk menyelesaikan masalah yang efisien. Cara kerja dari SMO ini adalah dengan mengoptimalkan hanya dua variabel *lagrange* dalam setiap iterasi.Proses untuk menemukan hyperplane menggunakan SVM menurut Campbell dirinci sebagai berikut (Sirait *et al*., 2019):

1. Terdapat data $\overrightarrow{X_1}\ \epsilon(\overrightarrow{X_1}, \overrightarrow{X_2}, \dots, \overrightarrow{X_n}$ dengan X_i merupakan data yang terdiri dari M atribut dan kelas target $y_i\ \epsilon$ +1, -1.
2. Kelas +1 dan -1 dipisahkan oleh *hyperplane* dengan menggunakan persamaan (2.1) sebagai berikut:

$$w.x + b = 0 \tag{2.1}$$

3. SVM bertujuan untuk menemukan *hyperplane* pemisah dengan memaksimalkan jarak antara dua kelas. Memaksimalkan jarak dengan mencari titik minimal persamaan:

$$min_w \frac{1}{2}(||w||)^2 \tag{2.3}$$

dengan kendala persamaan (2.3):

$$y_i(x_i w + b) - 1\ \geq 0 \tag{2.4}$$

4. Dengan membuat fungsi keputusan untuk persamaan sebagai berikut:

$$Linear: f(x_d) = \sum_{i=1}^{n} \alpha_i y_i (x_i . x_d) + b \tag{2.5}$$

dengan n merupakan jumlah *support vector*, $x_i$merupakan *support vector* ke- i, x_d merupakan data uji, $\alpha_i$merupakan bobot, $y_i$sebagai target ke- i, dan b adalah bias atau posisi bidang relatif terhadap pusat koordinat.

Karena kesederhanaan dan fleksibilitas dari SVM yang relatif untuk mengatasi berbagai masalah dalam klasifikasi, maka SVM secara khusus memberikan kinerja prediksi yang seimbang, bahkan dalam ukuran sampel yang mungkin terbatas. Oleh karena itu, peneliti menggunakan metode SVM dikarenakan kesederhanaan dan fleksibilitasnya.

2.5 *Synthetic Minority Oversampling Technique* (SMOTE)

Minority Overslamping Technique (SMOTE) adalah teknik menyeimbangkan distribusi data sampel pada kelas minoritas dengan cara menyeleksi data sampel agar jumlah data sampe menjadi seimbang dengan jumlah sampel pada kelas mayoritas (Kasanah *et al*., 2019). Teknik ini bekerja dengan cara mengidentifikasi data mayoritas yang terdekat dengan data minoritas, dan membuat data baru yang sama dengan data minoritas sintetis. Ini dapat membantu memperbaiki ketidakseimbangan klasifikasi data dan model klasifikasi. Dalam penggunaan SMOTE ini memungkinkan adanya *overvitting*. Berikut merupakan tahapan dalam melakukan SMOTE:

1. Menghitung jarak antar data pada data minoritas
2. Menetukan nilai presentase SMOTE yang akan digunakan untuk menciptakan data sintesis
3. Menciptakan data sintesis dengan persamaan sebagai berikut:

$$x_{syn} = x_i + (x_{knn} - x_i) \times y \tag{2.9}$$

dimana :
x_{syn} = data sintesis
x_i = data ke- i dari kelas minor
x_{knn} = data dari kelas minor yang memiliki jarak terdekat dari x_i
y = nilai random antara 0 dan 1.

2.6 *Principal Component Analysis (PCA)*

Principal Component Analysis (PCA) adalah suatu metode yang mampu mereduksi dimensi data yang terdiri dari variabel yang saling berhubungan dalam skala yang besar, yang menyimpan variabel sebanyak mungkin dari dataset tersebut (Ihsani *et al*., 2020). Dimensi data dikurangi dengan memproyeksikan ke dimensi yang lebih rendah. Komponen utama yang pertama dipilih untuk mengurangi jarak antara data dan proyeksi. Komponen selanjutnya juga dipilih dengan cara yang sama tetapi tidak berkorelasi dengan komponen utama sebelumnya. Akibatnya, dimensi data dikurangi dengan menghilangkan komponen yang lebih lemah dan menghilangkan informasi yang berlebihan (Hoq *et al*., 2021).

PCA memiliki tujuan untuk mengidentifikasi pola dan struktur yang mendasari data berdimensi tinggi dengan mengubahnya menjadi ruang berdimensi rendah dan juga mempertahankan informasi yang paling penting. Cara kerja dari PCA sendiri dengan menghilangkan korelasi antara variabel bebas melalui transformasi variabel bebas asal ke variabel baru yang tidak berkorelasi sama sekali, yang sering disebut dengan *principal component*. PCA ini merupakan metode pengembangan dari teori *Karhunen-Loève Transform* (KLT) yang merupakan suatu metode transformasi linear pada metode pengolahan citra. Penentuan jumlah PC berdasarkan pada *cumulative explained rasio* yang didasarkan dari komponen-komponen yang dapat dipertahankan dengan nilai variansnya minimal 95% sampai 1 (Jolliffe, 2019). Pada penelitian yang dilakukan Almais *et al* (2023) untuk menentukan jumlah PC dilihat dari *eigenvalue* dan varians rasionya (Almais *et al*., 2023). Lalu pada penelitian Purnama (2019),

memberikan beberapa cara untuk menentukan komponen utama, diantaranya; menggunakan nilai eigen >1; menggunakan proporsi kumulatif varian terhadap total varian dengan proporsi kumulatif varian oleh komponen utama minimal 80% dan proporsi total variansi populasi bernilai cukup besar (Purnama, 2019). Kemudian pada penelitian N. P. Sari *et al* (2023), penentuan konsep untuk komponen utama yang akan dipertahankan adalah yang memiliki nilai variasi >1 yang mana metode PCA ini mampu mengekstrak data dengan variabel menjadi jumlah yang lebih kecil tanpa menghilangkan informasi dari keseluruhan data. Syarat untuk menentukan PC terbaik adalah dengan memiliki nilai *cumulative proportion* diatas 80% (N. P. Sari *et al*., 2023).

Adapun tahap-tahap untuk melakukan reduksi dimensi menggunakan PCA (Sirait et al., 2019):

1. Data diatur sebagai satu set, n data vektor X_1…..X_n yang mana matriks *input* untuk PCA adalah $X_{(i,j)}$untuk *training*, dengan *i* merupakan baris dan *j* merupakan kolom.
2. Menghitung nilai mean dari data menggunakan rumus (2.6):

$$\bar{X} = \sum_{i=1}^{n} \frac{1}{X_i} \tag{2.6}$$

dimana:
$\bar{X}$ = *Mean*
n = Jumlah dari data observasi
X_i = Dimensi data ke i

3. Mencari matriks kovarians dari data menggunakan rumus (2.7):

$$C_X = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})\ (X_i - \bar{X})^T \quad (2.7)$$

dimana:
n = jumlah data observasi
X_i = data observasi
$\bar{X}$ = nilai rata-rata data atau *mean*

4. Menghitung nilai eigen (v_m) dan vektor eigen λ_m menggunakan rumus (2.8):

$$C_X v_m = \lambda_m v_m \quad (2.8)$$

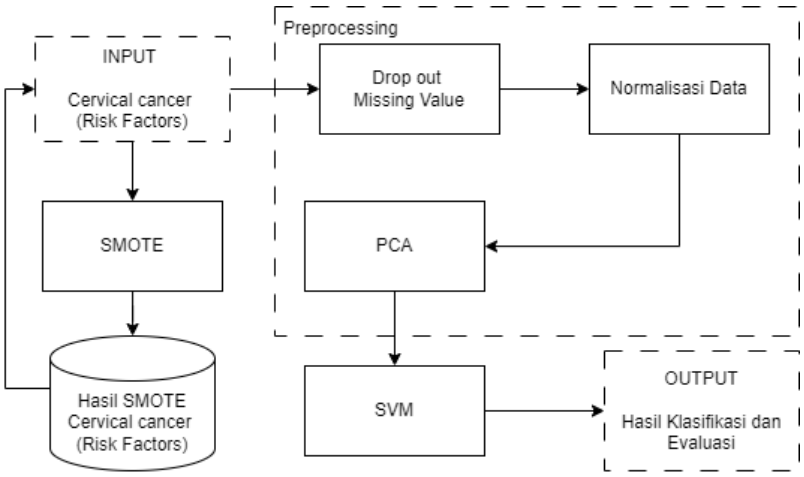
dimana:
C = Kovarian matriks
v = vektor eigen
λ = nilai eigen

5. Mengurutkan nilai eigen dalam urutan menurun
6. *Principal Component* (PC) adalah kumpulan vektor eigen sesuai dengan nilai eigen yang telah diurutkan dalam langkah 5.
7. Dimensi PC akan dikurangi berdasarkan pada nilai eigen.
8. Setelah itu pilih beberapa vektor eigen dengan nilai eigen tertinggi.
9. Selanjutnya, lakukan transformasi data menggunakan vektor eigen yang telah dipilih.

BAB III
DESAIN DAN IMPLEMENTASI

3.1 Desain Sistem

Untuk menyelesaikan masalah yang telah dibahas sebelumnya, tahapan perancangan sistem ini mencakup alur sistem dari input data hingga output yang sesuai dengan tujuan penelitian.. Berikut merupakan rancangan sistem yang akan dilakukan pada penelitian ini:



Gambar 3. 1 Desain Sistem

3.1.1 *Preprocessing*

Preprocessing adalah proses yang mana data akan dipersiapkan dan diperbaiki agar dapat digunakan untuk suatu pemodelan. Tujuan dari *preprocessing data* adalah untuk memastikan bahwa data yang akan diproses tersebut dalam keadaan yang bersih dan siap untuk dilakukan analisis. Pada *preprocessing data* ini akan melibatkan beberapa tahapan. Berikut merupakan tahapan-tahapan yang akan dilakukan pada praproses data:

3.1.1.1 *Dropout Missing Value*

Dropout Missing value atau nilai yang hilang adalah kondisi ketika dalam kumpulan data terdapat informasi yang semestinya ada, namun terdapat beberapa bagian dari data yang tidak ada atau kosong. Karena dalam penelitian menggunakan *dataset public*, jadi untuk menghindari data yang tidak memiliki nilai dan memastikan bahwa data yang digunakan pada penelitian ini disusun dengan baik, diperlukan *dropout missing value.*

3.1.1.2 Normalisasi Data

Normalisasi data adalah proses pengubahan skala data atau rentang nilai dalam dataset agar semua atribut atau fitur data memiliki skala yang sebanding. Tujuan dari adanya normalisasi ini adalah untuk menghilangkan perbedaan skala antara atribut-atribut dalam dataset sehingga setiap atribut memiliki pengaruh yang seimbang dalam analisis pemodelan data. Peneliti akan menggunakan *Min Max Scaller* untuk menskalakan nilai atribut dengan rentang 0-1. Rumus umum untuk normalisasi data terdapat pada persamaan (3.1):

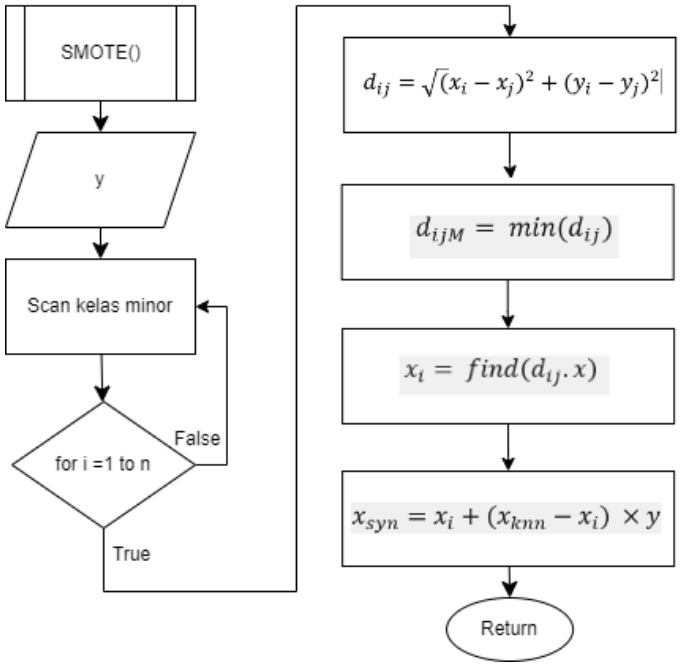
$$Normalisasi = \frac{data - min(data)}{\max(data) - \min(data)} \tag{3.1}$$

3.1.1.3 *Synthetic Minority Oversampling Technique (SMOTE)*

Data target penelitian ini adalah data yang tidak seimbang, sehingga perlu dilakukan teknik SMOTE agar distribusi dari kelas target menjadi seimbang. Persebaran kelas pada dataset *Cervical cancer (Risk Factors)* tiap targetnya tidak seimbang. Kelas 0 (negatif kanker serviks) pada target *Hinselmann* memiliki 823 dan kelas 1 (positif kanker serviks) pada target *Hinselmann* memiliki 35. Kelas 0

(negatif kanker serviks) pada target *Schiller* memiliki 784 dan kelas 1 (positif kanker serviks) pada target *Schiller* memiliki 74. Kelas 0 (negatif kanker serviks) pada target *Cytology* memiliki 814 dan kelas 1 (positif kanker serviks) pada target *Cytology* memiliki 44. Kelas 0 (negatif kanker serviks) pada target *Biopsy* memiliki 803 dan Kelas 1 (positif kanker serviks) pada target *Biopsy* memiliki 55. Dari 4 target tersebut persebaran kelas tidak seimbang, oleh karena itu digunakan teknik SMOTE untuk menangani ketidakseimbangan data.

Berikut adalah alur sistem *Synthetic Minority Oversampling Technique* (SMOTE) yang dibuat gambaran secara umum:



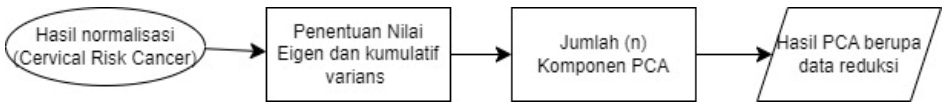
Gambar 3. 2 Alur algoritma SMOTE

Pada tahap SMOTE yang ditunjukkan pada gambar 3.3 dengan contoh simulasinya terdapat 1 kelas minoritas misalnya y =1 dan akan dilakukan replikasi untuk mencari tetangga terdekat x_{knn} dengan menggunakan jarak *Euclidean* untuk setiap

data dalam kelas tersebut. Data *synthetics* dari kelas minoritas akan dibangkitkan dari perhitungan jarak *Euclidean* semisal mendapatkan paling dekat $\sqrt{11}$ maka kelas akan dilakukan replikasi sebanyak lima kali. Yang awal data kelasnya berjumlah 1, maka jumlah datanya akan menjadi enam. Kemudian akan dihitung data sintetisnya. Dengan nilai y pada setiap data sintesis akan berbeda karena merupakan nilai acak antara 0 dan 1. Dengan data sintesis yang sudah dihitung maka akan mendapatkan lima data sintesis baru diantara data ke-1 dan data ke-2.

3.1.1.4 *Principal Component Analysis* (PCA)

Alur sistem Principal Component Analysis digambarkan sebagai berikut:



Gambar 3. 3 Proses Reduksi Dimensi PCA

Pada gambar 3.2 menjelaskan bahwa untuk mengurangi dimensi besar pada data, maka diperlukan sebuah teknik reduksi dimensi agar dapat meminimumkan dan *running time* ketika proses klasifikasi agar menjadi lebih baik. Pada hasil normalisasi akan dicari nilai eigen dengan menghitung korelasi antar data uji dari matriks kovarians yang melalui proses vector eigen sehingga akan membentuk matriks vektor eigen. Setelah itu akan dilakukan proses menghitung kumulatif varians untuk menentukan berapa komponen yang akan digunakan. Kemudian jumlah (n) akan diambil dari kumulatif diatas 95% atau varians yang berada diatas 0 dan hasilnya akan mendapatkan data yang sudah direduksi.

3.1.1.5 *Split Data*

Pembagian data atau yang sering disebut dengan "*split data*", memiliki peran penting dalam memisahkan dataset yang terdiri dari dua komponen utama: *training data* dan *testing data*. Dalam hal pemodelan data pada pembelajaran mesin, tindakan ini sangat penting.. *Training data* berguna untuk memodel data, sementara *testing data* berguna untuk menilai seberapa baik performa model yang sudah dibuat. Oleh sebab itu, *split data* ini dilakukan agar mendapatkan hasil yang akurat dari model yang dibuat. Penelitian akan membagi data menjadi empat bagian. Model A akan memiliki perbandingan 90% pelatihan dan 10% pengujian data, Model B akan memiliki perbandingan 80% pelatihan dan 20% pengujian data, Model C akan memiliki perbandingan 70% pelatihan dan 30% pengujian data, dan Model D akan memiliki perbandingan 60% pelatihan dan 40% pengujian data.

3.2 Pengumpulan Data

Setelah menentukan topik penelitian, langkah berikutnya adalah mencari data yang diperlukan untuk melakukan penelitian. Pada penelitian ini data diambil dari *UCI Machine Learning Repository: Cervical cancer (Risk Factors)*.

Data ini merupakan open access data web *UCI Machine Learning* (https://archive.ics.uci.edu/dataset/383/cervical+cancer+risk+factors). Data ini dikumpulkan dari Rumah Sakit Universitario de Caracas' di Caracas, Venezuela. Kumpulan data ini terdiri dari informasi demografis, kebiasaan, dan riwayat medis 858 pasien yaitu 36 label yaitu 32 faktor resiko dan 4 target, yaitu *Hinselmann*, *Schiller*, *Cytology*, dan *Biopsy*. Namun terdapat beberapa pasien yang memutuskan untuk tidak menjawab beberapa pertanyaan karena terdapat masalah privasi (nilai

yang hilang). Data ini disumbangkan pada tanggal 3 Februari 2017. Penjelasan atribut-atribut pada dataset dapat dilihat pada tabel berikut.

Tabel 3. 1 Deskripsi Data

No	Atribut	Penjelasan
1.	*Age*	Berapa usia pasien saat data diambil
2.	*Number of sexual partners*	Jumlah pasangan seksual yang pernah dimiliki pasien
3.	*First sexual intercourse*	Usia pertama kali melakukan hubungan seksual
4.	*Num of pregnancies*	Berapa kali pasien hamil
5.	*Smokes*	0 jika seseorang tidak merokok, 1 jika seseorang pernah merokok
6.	*Smokes (years)*	Jumlah tahun seseorang telah merokok
7.	*Smokes (packs/year)*	Jumlah bungkus rokok yang dihisap seseorang per tahun
8.	*Hormonal Contraceptives*	0 jika seseorang tidak menggunakan kontrasepsi hormonal, 1 jika seseorang menggunakan kontrasepsi hormonal
9.	*Hormonal Contraceptives (years)*	Jumlah tahun seseorang menggunakan kontrasepsi hormonal
10.	*IUD*	0 jika seseorang tidak memiliki IUD, 1 jika seseorang memiliki IUD
11.	*IUD(years)*	Jumlah tahun seseorang telah memakai IUD
12.	*STDs*	0 jika seseorang tidak mengidap STDs, 1 jika seseorang memang mengidap STDs
13.	*STDs (number)*	Jumlah STDs yang diderita seseorang
14.	*STDs:condylomatosis*	0 jika seseorang tidak menderita kondilomatosis, 1 jika seseorang memang menderita kondilomatosis
15.	*STDs:cervical condylomatosis*	0 jika seseorang tidak menderita kondilomatosis serviks, 1 jika seseorang memang menderita kondilomatosis serviks
16.	*STDs:vaginal condylomatosis*	0 jika seseorang tidak menderita kondilomatosis vagina, 1 jika seseorang memang menderita kondilomatosis vagina
17.	*STDs:vulvo-perineal condylomatosis*	0 jika seseorang tidak menderita kondilomatosis vulvo-perineum, 1 jika seseorang memang menderita kondilomatosis vulvo-perineum
18.	*STDs:syphilis*	0 jika seseorang tidak mengidap sifilis, 1 jika seseorang mengidap sifilis
19.	*STDs:pelvic inflammatory disease*	0 jika seseorang tidak mengidap PID, 1 jika seseorang memang mengidap PID
20.	*STDs:genital herpes*	0 jika seseorang tidak menderita herpes genital, 1 jika seseorang memang menderita herpes genital
21.	*STDs:molluscum contagiosum*	0 jika seseorang tidak menderita moluskum kontagiosum, 1 jika seseorang menderita moluskum kontagiosum
22.	*STDs:AIDS*	0 jika seseorang tidak mengidap AIDS, 1 jika seseorang memang mengidap AIDS
23.	*STDs:HIV*	0 jika seseorang tidak mengidap HIV, 1 jika seseorang memang mengidap HIV
24.	*STDs:Hepatitis B*	0 jika seseorang tidak mengidap Hepatitis B, 1 jika seseorang mengidap Hepatitis B
25.	*STDs:HPV*	0 jika seseorang tidak mengidap HPV, 1 jika seseorang mengidap HPV
26.	*STDs: Number of diagnosis*	Jumlah STDs yang didiagnosis

No	Atribut	Penjelasan
27.	*STDs: Time since first diagnosis*	Jumlah tahun yang telah berlalu sejak diagnosis STDs pertama
28.	*STDs: Time since last diagnosis*	Jumlah tahun yang telah berlalu sejak diagnosis STDs terakhir
29.	*Dx:Cancer*	0 jika sebelumnya tidak ada diagnosis kanker, 1 jika sebelumnya ada diagnosis
30.	*Dx:CIN*	0 jika sebelumnya tidak ada diagnosis neoplasia intraepitel serviks, 1 jika ada diagnosis sebelumnya
31.	*Dx:HPV*	0 jika sebelumnya tidak ada diagnosis HPV, 1 jika sebelumnya ada diagnosis
31.	*Dx:HPV*	0 jika sebelumnya tidak ada diagnosis HPV, 1 jika sebelumnya ada diagnosis
32.	*Dx*	0 jika sebelumnya tidak ada diagnosis kanker serviks, 1 jika sebelumnya ada diagnosis
33.	*Hinselmann*	0 menunjukkan tidak ada kanker, 1 menunjukkan kanker
34.	*Schiller*	0 menunjukkan tidak ada kanker, 1 menunjukkan kanker
35.	*Cytology*	0 menunjukkan tidak ada kanker, 1 menunjukkan kanker
36.	*Biopsy*	0 menunjukkan tidak ada kanker, 1 menunjukkan kanker

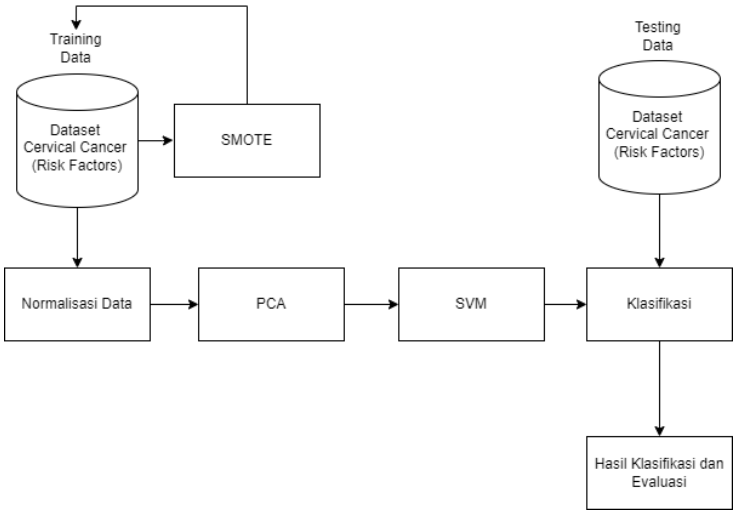
Tabel 3. 2 Contoh Dataset

Age	*First sexual intercourse*	*Number of Pregnancies*	*Smokes*	*Hormonal Contraceptives*	*Dx: HPV STD: genital herpes*	*Hinselmann*	*Schiller*	*Cytology*	*Biopsy*
18	15	4	15	0	0	0	0	0	0
42	21	3	21	0	1	0	0	0	1
40	18	1	18	1	0	0	0	0	1
40	17	1	17	0	0	0	0	0	1
37	26	6	26	0	0	0	0	0	0
38	15	2	15	0	0	0	0	0	0
52	16.0	5.0	16.0	0	0	1	1	0	0
46	21.0	3.0	21.0	0	0	0	0	0	0
42	23.0	3.0	23.0	0	0	0	0	0	0
51	17.0	3.0	17.0	0	0	0	0	0	1
26	26.0	1.0	26.0	0	0	0	0	1	0
45	20.0	1.0	20.0	0	1	1	0	0	0

3.3 Implementasi *Support Vector Machine* (SVM)

Setelah Setelah dilakukan reduksi dimensi PCA dan sudah melewati tahap-tahap *preprocessing*, hasil dari dataset yang sudah dinormalisasikan dan distandarisasi akan masuk pada implementasi SVM. Setelah dataset tersebut diolah dengan PCA untuk menemukan parameter yang cocok, kemudian akan dibagi

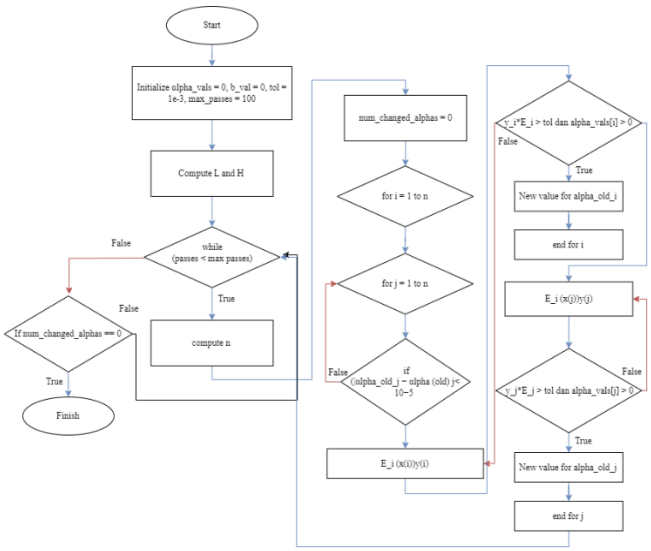
menjadi dua yaitu *training data* dan *testing data*. Dan tahap selanjutnya yaitu memasukkan *training data* ke implementasi SVM dan akan membentuk model SVM. Setelah itu akan dilakukan pengujian dengan menggunakan *testing data* yang akan digunakan untuk menganalisis seberapa akurasi yang didapat pada model yang sudah dibangun tersebut. Berikut merupakan alur pada proses training dan testing dalam klasifikasi kanker serviks berrdasarkan metode SVM berbasis PCA.



Gambar 3. 4 Proses *Training* dan *Testing*

Pada gambar 3.4 menggambarkan alur proses pelatizhan dan pengujian model SVM untuk klasifikasi risiko kanker serviks. Proses ini dimulai dengan *spit data*, yang mana dataset *Cervical Cancer* digunakan sebagai input. Data tersebut akan mengalami proses normalisasi untuk memastikan bahwa semua fitur berada ada skala yang sama. Kemudian akan diterapkan PCA untuk mengurangi dimensi data, sehingga hanya fitur-fitur tertentu yang paling penting yang dipertahankan. Setelah itu, akan masuk ke SMOTE yang digunakan untuk menangani ketidakseimbangan kelas dalam dataset dengan cara menghasilkn data sampe; minoritas sintesis. Kemudian data yang sudah diproses akan digunakan untuk

pelatihan model SVM. Dan pada proses pelatihan menghasilkan bobot-bobot yang disimpan dan akan digunakan pada tahap pengujian. Data pengujian akan melalui proses klasifikasi menggunakan model SVM yang sudah dilatih sebelumnya dengan bobot yang sudah ditentukan. Kemudian hasil klasifikasi akan dievaluasi untuk mengukur kinerja dari sistem ini. Pada proses implementasi SVM , berikut adalah alur yang digunakan.



Gambar 3. 5 *Flowchart Support Vector Machine* (SVM)

Pada gambar 3.5 menggambarkan proses *Sequential Minimal Optimization* (SMO) dalam *Support Vector Machine* (SVM). Langkah pertama setelah data dimasukkan maka akan dilakukan inisialisasi *alphas_vas, b_val, tol (toleransi), max_passes*. Setelah itu dilakukan pengecekan untuk batas bawah (L) dan batas atas (H). Kemudian akan masuk pada pengulangan selama jumlah iterasi (*passes*) masih kurang dari jumlah maksimum iterasi (*max_passes*), maka program akan masuk ke dalam *loop `while`*. Dengan menghitung jumlah vektor pendukung yang memiliki alpha lebih besar dari 0 yang mana jumlah perubahan alpha diatur menjadi 0. Kemudian akan dilakukan iterasi setiap sampel dan sampel lain dalam dataset. Dan

akan masuk kondisi jika alpha_old_j - alpha_old_i < 10^-5 terpenuhi, maka nilai alpha untuk sampel j diperbarui. Kemudian akan menghitung nilai eror yang digunakan untuk memperbarui nilai alpha. Dan setelah nilai terpenuhi maka nilai alpha untuk sampel i akan diperbarui. Setelah selesai iterasi untuk semua sampel dalam dataset akan kembali ke langkah *loop* i dan j. Kemudian akan dilakukan pengecekan terhadap kondisi apakah ada peerubahan pada nilai alpha. Jika jumlah iterasi telah mencapai *max_passes* atau tidak terjadi perubahan pada nilai alpha, maka akan selesai prosesnya.

3.4 Evaluasi Performa

Evaluasi performa merupakan metode yang digunakan untuk mengukur seberapa baik suatu model klasifIkasi seperti halnya pada *machine learning* untuk memprediksi label atau kelas yang benar dari data. Pada penelitian ini akan menggunakan *confusion matrix* untuk mengukur performa model klasifikasi.

Confusion Matrix adalah sebuah alat ukur dalam bentuk matrix yang digunakan untuk menghitung sejauh mana kelas-kelas dalam algoritma yang digunakan dalam ketepatan klasifikasi.

Tabel 3. 3 *Confusion Matrix*

Confusion Matrix		Nilai Prediksi	
		True	False
Nilai Sebenarnya	*TRUE*	TP (*True Positive Correct result*)	FN (*False Negative Correct result*)
	FALSE	FP (*False Positive Correct result*)	TN (*True Negative Correct result*)

Confusion Matrix digunakan untuk mengevaluasi kinerja dari model *Support Vector Machine*. Langkah pertama yang dilakukan adalah pembagian data, setelah itu melatih model *Support Vector Machine*. Kemudian setelah melakukan pelatihan tersebut, jumlah prediksi yang benar dan yang salah dari model *Support Vector Machine* untuk setiap kelas ditunjukkan dengan *confusion matrix* dibandingkan dengan label kelas data uji yang sebenarnya.. Berikut merupakan rumus *accuracy, precision, recall, dan f1 score* :

$$Accuracy = \frac{\mathrm{TP} + \mathrm{TN}}{TP + FP + TN + FN} \tag{3.2}$$

accuracy merupakan presentase dari total data yang terprediksi dengan benar.

$$Precision = \frac{\mathrm{TP}}{TP + FP} \tag{3.3}$$

precision merupakan nilai ketepatan prediksi dengan benar. Perhitungan *precision* ini dilakukan dengan membandingkan jumlah prediksi benar dengan jumlah seluruh prediksi.

$$Recall = \frac{\mathrm{TP}}{TP + FN} \tag{3.4}$$

recall merupakan nilai perbandingan ketepatan prediksi benar dengan jumlah seluruh prediksi yang seharusnya benar.

$$f1\ score = 2\frac{(\mathrm{recall} * \mathrm{precision})}{(recall + precision)} \tag{3.5}$$

f1 score merupakan perhitungan yang digunakan untuk menggabungkan nilai *precision* dan *recall*.

3.5 Skenario Pengujian

Berikut merupakan skenario pengujian yang akan dilakukan oleh peneliti dengan rasio perbandingan data latih dibanding dengan data uji.

Tabel 3. 4 Skenario Pengujian

Skenario	Rasio Perbandingan	Target
Pengujian dengan Metode SVM PCA SMOTE	90:10	*Hinselmann, Schiller, Cytology, Biopsy.*
	80:20	
	70:30	
	60:40	
Pengujian dengan Metode SVM SMOTE tanpa PCA	90:10	
	80:20	
	70:30	
	60:40	
Pengujian dengan Metode SVM PCA tanpa SMOTE	90:10	
	80:20	
	70:30	
	60:40	
Pengujian dengan Metode SVM tanpa PCA SMOTE	90:10	
	80:20	
	70:30	
	60:40	

Pada tabel 3.4 dilakukan pengujian dengan rasio perbandingan data latih dan data uji 90:10 pada setiap target, pengujian dengan rasio perbandingan data latih dan data uji 80:20 pada setiap target, pengujian dengan rasio perbandingan data latih dan data uji 70:30 pada setiap target, dan pengujian dengan rasio perbandingan data latih dan data uji 60:40 pada setiap target.

BAB IV
HASIL DAN PEMBAHASAN

Bab ini memberikan penjelasan tentang hasil penelitian yang telah dilakukan dan membahas hasil perhitungan *Principal Component Analysis* (PCA). PCA memberikan analisis fitur yang dominan dari komponen utama yang sesuai dan perhitungan *Support Vector Machine* (SVM) untuk mengklasifikasikan kanker serviks berdasarkan hasil PCA sesuai dengan skenario pengujian.

4.1 Hasil SMOTE (*Synthetic Minority Over-sampling Technique*)

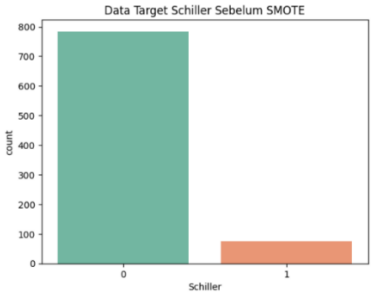
Uji coba pada penelitian ini akan dilakukan agar mendapatkan tingkat akurasi paling baik dari 4 target pada dataset yaitu *Hinselmann, Schiller, Cytology*, dan *Biopsy*. Untuk memastikan keseimbangan pada nilai target, maka peneliti akan menggunakan SMOTE untuk meningkatkan kelas minoritas sehingga setiap kelas memiliki jumlah data yang seimbang. Dengan adanya SMOTE, sampel dari kelas minoritas diperbanyak secara sintetis agar mendapatkan keseimbangan data dengan kelas mayoritas.

Proses ini dapat dilakukan dengan menghasilkan sampel baru berdasarkan interpolasi antara sampel minoritas yang ada dan tetangga terdekatnya. Ketika SMOTE digunakan, target dari sampel minoritas akan diperbanyak dengan menciptakan sampel baru, maka nilai variabel (fitur) dari sampel baru juga akan di SMOTE sehingga fitur-fitur tersebut mencerminkan distribusi yang seimbang anatra kelas-kelas yang berbeda. Dan pada hasil uji coba akan diberikan hasil

dengan menggunakan PCA dan hasil dengan tidak menggunakan PCA dan juga pada metode SVM akan menggunakan random hyperparameter terbaik.

Gambar 4. 1 Data Target *Hinselmann* sebelum SMOTE

Pada gambar 4.1 menunjukkan bahwa target pada penelitian ini yaitu *Hinselmann* terdapat *imbalanced* data (data tidak seimbang). Pada target *Hinselmann*, kelas 0 (negatif kanker serviks) memiliki 823 data dan kelas 1 (positif kanker serviks) memiliki 35 data.

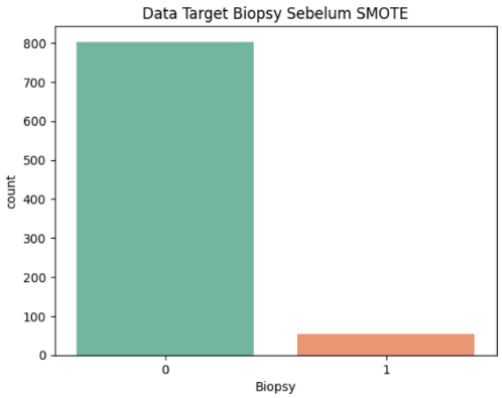


Gambar 4. 2 Data Target *Schiller* sebelum SMOTE

Pada gambar 4.2 menunjukkan bahwa target pada penelitian ini yaitu *Schiller* terdapat *imbalanced* data (data tidak seimbang). Pada target *Schiller*, kelas 0 memiliki 784 data dan kelas 1 memiliki 74 data.

Gambar 4. 3 Data Target *Cytology* sebelum SMOTE

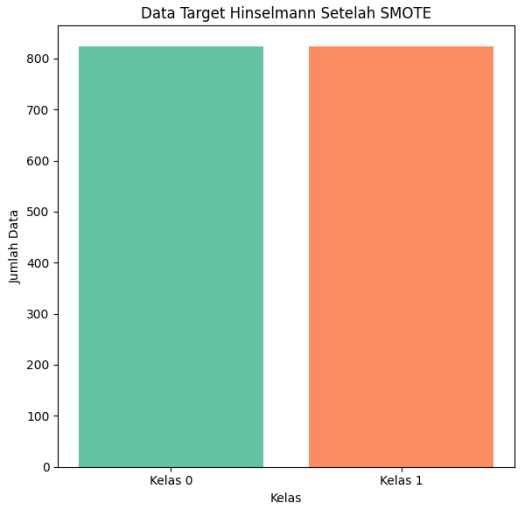
Pada gambar 4.3 menunjukkan bahwa target pada penelitian ini yaitu *Cytology* terdapat *imbalanced* data (data tidak seimbang). Pada target *Cytology*, kelas 0 memiliki 814 data dan kelas 1 memiliki 44 data.



Gambar 4. 4 Data Target *Biopsy* sebelum SMOTE

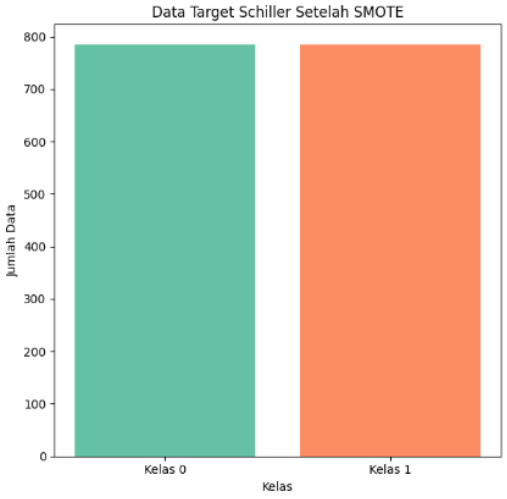
Pada gambar 4.4 menunjukkan bahwa target pada penelitian ini yaitu *Biopsy* terdapat *imbalanced* data (data tidak seimbang). Pada target *Biopsy*, kelas 0 memiliki 803 data dan kelas 1 memiliki 55 data.

Dan berikut merupakan visualisasi data setelah dilakukan teknik *Synthetic Minority Oversampling Technique* (SMOTE).



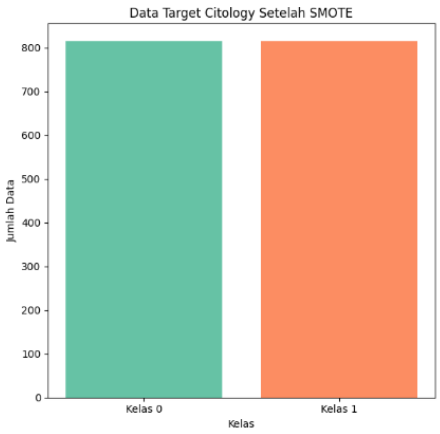
Gambar 4. 5 Data Target *Hinselmann* sesudah SMOTE

Pada gambar 4.5 menunjukkan bahwa dari target *Hinselmann* sudah seimbang antar kelasnya, yang awalnya kelas 1 memiliki 35 data setelah dilakukan teknik SMOTE menjadi 823 data.



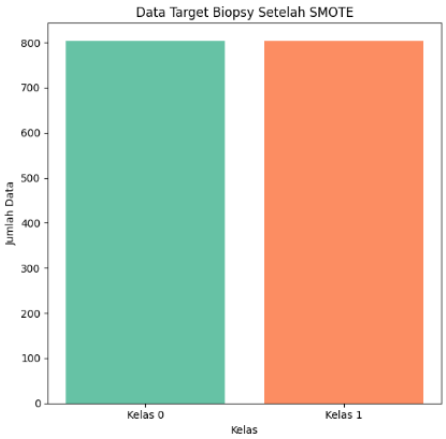
Gambar 4. 6 Data Target *Schiller* sesudah SMOTE

Pada gambar 4.6 menunjukkan bahwa dari target *Schiller* sudah seimbang antar kelasnya, yang awalnya kelas 1 memiliki 74 data setelah dilakukan teknik SMOTE menjadi 784 data.



Gambar 4. 7 Data Target *Cytology* sesudah SMOTE

Pada gambar 4.7 menunjukkan bahwa dari target *Schiller* sudah seimbang antar kelasnya, yang awalnya kelas 1 memiliki 44 data setelah dilakukan teknik SMOTE menjadi 814 data.



Gambar 4. 8 Data Target *Biopsy* sesudah SMOTE

Pada gambar 4.8 menunjukkan bahwa dari target *Biopsy* sudah seimbang antar kelasnya, yang awalnya kelas 1 memiliki 55 data setelah dilakukan teknik SMOTE menjadi 803 data. Kegunaan SMOTE ini dapat meningkatkan nilai sensitivitas lebih dari 50% walaupun nilai akurasi dan spesifitasnya menurun dibandingkan dengan pemodelan sebelum SMOTE (Dwi Astuti & Nova Lenti, 2021).

4.2 Hasil *Principal Component Analysis* (PCA)

Untuk mengurangi ukuran variabel dalam kumpulan data, perlu digunakan teknik *Principal Component Analysis* (PCA). Tujuannya adalah untuk menyederhanakan dan menghilangkan faktor atau indikator skrining yang kurang dominan dan kurang relevan tanpa mengurangi maksud dan tujuan dari data asli *Cervical cancer (Risk Factors) dataset*. Sebelum melakukan teknik PCA, terlebih dahulu akan dilakukan teknik SMOTE. Terdapat empat target dalam penelitian ini: *Hinselmann, Schiller, Cytology,* dan *Biopsy*. Setiap target memiliki distribusi data positif dan negatif yang berbeda, jadi penerapan SMOTE bertujuan untuk menyeimbangkan jumlah sampel pada masing-masing kelas. Setelah proses penyeimbangan data, PCA digunakan untuk mengurangi dimensi variabel dalam kumpulan data. Hasil komponen utama PCA tidak sama karena distribusi awal data positif dan negatif tidak sama untuk setiap target.

4.2.1 Hasil PCA Target *Hinselmann* setelah SMOTE

Berikut merupakan hasil perhitungan dari dataset *Cervical cancer (Risk Factors)* dengan target *Hinselmann* pada tabel 4.1 diketahui dengan rincian tiap hasil PC. Berikut tabel dari perhitungan PCA berdasarkan nilai eigen yang didapatkan.

Tabel 4. 1 *Eigen Value* dan Variance Rasio dari Target *Hinselmann*

Komponen Utama	***Eigenvalue***	***Varians Rasio (%)***	***Cumulative Variance Rasio***
PC1	6.419646e+00	23.57	0.235734
PC2	5.461148e+00	20.05	0.436272
PC3	2.694188e+00	9.89	0.535205
PC4	1.633427e+00	6.00	0.595185
PC5	1.341265e+00	4.93	0.644438
PC6	1.316729e+00	4.84	0.692789
PC7	8.764722e-01	3.22	0.724974
PC8	7.542803e-01	2.77	0.752672

Komponen Utama	***Eigenvalue***	***Varians Rasio (%)***	***Cumulative Variance Rasio***
PC9	7.202519e-01	2.64	0.779120
PC10	5.993066e-01	2.20	0.801127
PC11	5.768096e-01	2.12	0.822308
PC12	5.578332e-01	2.05	0.842792
PC13	5.290686e-01	1.94	0.862220
PC14	5.177036e-01	1.90	0.881230
PC15	5.105917e-01	1.87	0.899979
PC16	5.049237e-01	1.85	0.918521
PC17	4.876201e-01	1.79	0.936426
PC18	4.340388e-01	1.59	0.952365
PC19	3.784145e-01	1.39	0.966260
PC20	2.115201e-01	0.78	0.974027
PC21	1.741291e-01	0.64	0.980422
PC22	1.687315e-01	0.62	0.986618
PC23	1.282805e-01	0.47	0.991328
PC24	1.235374e-01	0.45	0.995864
PC25	5.579095e-02	0.20	0.997913
PC26	5.228565e-02	0.19	0.999833
PC27	4.543704e-03	0.02	1.000000
PC28	1.266052e-30	0.00	1.000000
PC29	3.047452e-32	0.00	1.000000
PC30	3.047452e-32	0.00	1.000000

Pada tabel 4.1 dapat dilihat bahwa nilai *variance rasio* yang lebih dari 1 dan *cumulative variance rasio*nya lebih dari 95% terdapat bahwa pada PC1 memiliki nilai *variance rasio* tertinggi yaitu 23.57 yang merupakan kontribusi tertinggi terhadap total varians data dan pada *cumulative variance rasio* juga akan meningkat di setiap PCnya. Pada PC1, *cumulative variance ratio* mencapai 0.235734 , yang berarti PC1 sendiri dapat menjelaskan sekitar 23.57% dari total varians data. an nilai variance rasio diatas 1 terdapat pada PC19 dengan nilai 1.39 dan*cumulative variance ratio* mencapai 0.966260, yang menunjukkan bahwa 19 komponen utama tersebut bersama-sama dapat menjelaskan sekitar 96.62% dari total varians data. Dengan demikian, grafik tersebut menggambarkan bahwa kontirbusi masing-masing PC terhadap total varians data dan meningkatnya kumulatif seiring dengan

penambahan PC. Hal ini menunjukkan berapa banyak PC yang akan dipertahankan dalam reduksi dimensi data.

4.2.2 Hasil PCA Target *Schiller* setelah SMOTE

Berikut merupakan hasil perhitungan dari dataset *Cervical cancer (Risk Factors)* dengan target *Schiller* pada tabel 4.2 diketahui dengan rincian tiap hasil PC. Berikut tabel dari perhitungan PCA berdasarkan nilai eigen yang didapatkan.

Tabel 4. 2 *Eigen Value* dan *Variance Rasio* dari Target *Schiller*

Komponen Utama	***Eigenvalue***	***Varians Rasio (%)***	***Cumulative Variance Rasio***
PC1	7.876932e+00	24.42	0.244164
PC2	6.366708e+00	19.74	0.441515
PC3	3.062923e+00	9.49	0.536458
PC4	2.164802e+00	6.71	0.603561
PC5	1.709949e+00	5.30	0.656565
PC6	1.560898e+00	4.84	0.704949
PC7	1.129293e+00	3.50	0.739954
PC8	9.056086e-01	2.81	0.768026
PC9	8.436757e-01	2.62	0.794177
PC10	7.375579e-01	2.29	0.817040
PC11	6.136214e-01	1.90	0.836060
PC12	6.019630e-01	1.87	0.854720
PC13	5.602665e-01	1.74	0.872086
PC14	5.456846e-01	1.69	0.889001
PC15	5.425471e-01	1.68	0.905819
PC16	5.354713e-01	1.66	0.922417
PC17	5.287118e-01	1.64	0.938806
PC18	5.121263e-01	1.59	0.954680
PC19	3.711229e-01	1.15	0.966184
PC20	2.798762e-01	0.87	0.974859
PC21	2.283539e-01	0.71	0.981938
PC22	1.988252e-01	0.62	0.988101
PC23	1.398836e-01	0.43	0.992437
PC24	1.117709e-01	0.35	0.995901
PC25	6.815687e-02	0.21	0.998014
PC26	5.927676e-02	0.18	0.999852
PC27	4.788468e-03	0.01	1.000000
PC28	2.784244e-30	0.00	1.000000
PC29	4.054923e-32	0.00	1.000000
PC30	4.054923e-32	0.00	1.000000

Pada tabel 4.2 dapat dilihat bahwa nilai *variance rasio* yang lebih dari 1 dan *cumulative variance rasio*nya lebih dari 95% terdapat bahwa pada PC1 memiliki nilai *variance rasio* tertinggi yaitu 24.42 yang merupakan kontribusi tertinggi terhadap total varians data dan pada *cumulative variance rasio* juga akan meningkat di setiap PCnya. Pada PC1, *cumulative variance ratio* mencapai 0.244164, yang berarti PC1 sendiri dapat menjelaskan sekitar 24.42% dari total varians data. Dan nilai variance rasio diatas 1 terdapat pada PC19 dengan nilai 1.15 dan *cumulative variance ratio* mencapai 0.966184, yang menunjukkan bahwa 19 komponen utama tersebut bersama-sama dapat menjelaskan sekitar 96.66% dari total varians data. Dengan demikian, grafik tersebut menggambarkan bahwa kontirbusi masing-masing PC terhadap total varians data dan meningkatnya kumulatif seiring dengan penambahan PC. Hal ini menunjukkan berapa banyak PC yang akan dipertahankan dalam reduksi dimensi data.

4.2.3 Hasil PCA Target *Citology* setelah SMOTE

Berikut merupakan hasil perhitungan dari dataset *Cervical cancer (Risk Factors)* dengan target *Citology* pada tabel 4.3 diketahui dengan rincian tiap hasil PC. Berikut tabel dari perhitungan PCA berdasarkan nilai eigen yang didapatkan.

Tabel 4.3 *Eigen Value* dan *Variance Rasio* dari Target *Citology*

Komponen Utama	***Eigenvalue***	***Varians Rasio (%)***	***Cumulative Variance Rasio***
PC1	7.876932e+00	24.42	0.244164
PC2	6.366708e+00	19.74	0.441515
PC3	3.062923e+00	9.49	0.536458
PC4	2.164802e+00	6.71	0.603561
PC5	1.709949e+00	5.30	0.656565
PC6	1.560898e+00	4.84	0.704949
PC7	1.129293e+00	3.50	0.739954
PC8	9.056086e-01	2.81	0.768026
PC9	8.436757e-01	2.62	0.794177
PC10	7.375579e-01	2.29	0.817040

Komponen Utama	***Eigenvalue***	***Varians Rasio (%)***	***Cumulative Variance Rasio***
PC11	6.136214e-01	1.90	0.836060
PC12	6.019630e-01	1.87	0.854720
PC13	5.602665e-01	1.74	0.872086
PC14	5.456846e-01	1.69	0.889001
PC15	5.425471e-01	1.68	0.905819
PC16	5.354713e-01	1.66	0.922417
PC17	5.287118e-01	1.64	0.938806
PC18	5.121263e-01	1.59	0.954680
PC19	3.711229e-01	1.15	0.966184
PC20	2.798762e-01	0.87	0.974859
PC21	2.283539e-01	0.71	0.981938
PC22	1.988252e-01	0.62	0.988101
PC23	1.398836e-01	0.43	0.992437
PC24	1.117709e-01	0.35	0.995901
PC25	6.815687e-02	0.21	0.998014
PC26	5.927676e-02	0.18	0.999852
PC27	4.788468e-03	0.01	1.000000
PC28	2.784244e-30	0.00	1.000000
PC29	4.054923e-32	0.00	1.000000
PC30	4.054923e-32	0.00	1.000000

Pada tabel 4.3 dapat dilihat bahwa nilai *variance rasio* yang lebih dari 1 dan *cumulative variance rasio*nya lebih dari 95% terdapat bahwa pada PC1 memiliki nilai *variance rasio* tertinggi yaitu 23.85 yang merupakan kontribusi tertinggi terhadap total varians data dan pada *cumulative variance rasio* juga akan meningkat di setiap PCnya. Pada PC1, *cumulative variance ratio* mencapai 0.238493, yang berarti PC1 sendiri dapat menjelaskan sekitar 23.85% dari total varians data. Dan nilai variance rasio diatas 1 terdapat pada PC19 dengan nilai 1.24 dan *cumulative variance ratio* mencapai 0.973608, yang menunjukkan bahwa 19 komponen utama tersebut bersama-sama dapat menjelaskan sekitar 97.36% dari total varians data. Dengan demikian, grafik tersebut menggambarkan bahwa kontirbusi masing-masing PC terhadap total varians data dan meningkatnya kumulatif seiring dengan penambahan PC. Hal ini menunjukkan berapa banyak PC yang akan dipertahankan dalam reduksi dimensi data.

4.2.4 Hasil PCA Target *Biopsy* setelah SMOTE

Berikut merupakan hasil perhitungan dari dataset *Cervical cancer (Risk Factors)* dengan target *Biopsy* pada tabel 4.4 diketahui dengan rincian tiap hasil PC. Berikut tabel dari perhitungan PCA berdasarkan nilai eigen yang didapatkan.

Tabel 4. 4 *Eigen Value* dan *Variance Rasio* dari Target *Biopsy*

Komponen Utama	***Eigenvalue***	***Varians Rasio (%)***	***Cumulative Variance Rasio***
PC1	7.252668e+00	21.81	0.218053
PC2	6.052921e+00	18.20	0.400036
PC3	2.867151e+00	8.62	0.486238
PC4	2.823451e+00	8.49	0.571125
PC5	2.049913e+00	6.16	0.632757
PC6	1.779477e+00	5.35	0.686257
PC7	1.533045e+00	4.61	0.732348
PC8	1.451952e+00	4.37	0.776002
PC9	9.035804e-01	2.72	0.803168
PC10	7.938391e-01	2.39	0.827035
PC11	6.858397e-01	2.06	0.847655
PC12	5.933158e-01	1.78	0.865493
PC13	5.759568e-01	1.73	0.882810
PC14	5.344957e-01	1.61	0.898879
PC15	5.321044e-01	1.60	0.914877
PC16	5.231928e-01	1.57	0.930607
PC17	5.141585e-01	1.55	0.946065
PC18	4.690880e-01	1.41	0.960169
PC19	3.601734e-01	1.08	0.970997
PC20	2.272089e-01	0.68	0.977829
PC21	1.899506e-01	0.57	0.983539
PC22	1.849257e-01	0.56	0.989099
PC23	1.505614e-01	0.45	0.993626
PC24	9.850040e-02	0.30	0.996587
PC25	5.767778e-02	0.17	0.998321
PC26	5.117104e-02	0.15	0.999860
PC27	4.658278e-03	0.01	1.000000
PC28	3.463249e-30	0.00	1.000000
PC29	3.733726e-32	0.00	1.000000
PC30	3.733726e-32	0.00	1.000000

Pada tabel 4.4 dapat dilihat bahwa nilai *variance rasio* yang lebih dari 1 dan *cumulative variance rasio*nya lebih dari 95% terdapat bahwa pada PC1 memiliki nilai *variance rasio* tertinggi yaitu 21.81 yang merupakan kontribusi tertinggi terhadap total varians data dan pada *cumulative variance rasio* juga akan meningkat

di setiap PCnya. Pada PC1, *cumulative variance ratio* mencapai 0.218053, yang berarti PC1 sendiri dapat menjelaskan sekitar 21.81% dari total varians data. Dan nilai variance rasio diatas 1 terdapat pada PC19 dengan nilai 1.08 dan *cumulative variance ratio* mencapai 0.970997, yang menunjukkan bahwa 19 komponen utama tersebut bersama-sama dapat menjelaskan sekitar 97.09% dari total varians data. Dengan demikian, grafik tersebut menggambarkan bahwa kontirbusi masing-masing PC terhadap total varians data dan meningkatnya kumulatif seiring dengan penambahan PC. Hal ini menunjukkan berapa banyak PC yang akan dipertahankan dalam reduksi dimensi data.

4.2.5 Hasil PCA tanpa SMOTE

Berikut merupakan hasil perhitungan dari dataset *Cervical cancer (Risk Factors)* pada tabel 4.5 diketahui dengan rincian tiap hasil PC. Berikut tabel dari perhitungan PCA berdasarkan nilai eigen yang didapatkan.

Tabel 4. 5 *Eigen Value* dan *Variance Rasio*

Komponen Utama	***Eigenvalue***	***Varians Rasio (%)***	***Cumulative Variance Rasio***
PC1	7.252668e+00	21.81	0.218053
PC2	6.052921e+00	18.20	0.400036
PC3	2.867151e+00	8.62	0.486238
PC4	2.823451e+00	8.49	0.571125
PC5	2.049913e+00	6.16	0.632757
PC6	1.779477e+00	5.35	0.686257
PC7	1.533045e+00	4.61	0.732348
PC8	1.451952e+00	4.37	0.776002
PC9	9.035804e-01	2.72	0.803168
PC10	7.938391e-01	2.39	0.827035
PC11	6.858397e-01	2.06	0.847655
PC12	5.933158e-01	1.78	0.865493
PC13	5.759568e-01	1.73	0.882810
PC14	5.344957e-01	1.61	0.898879
PC15	5.321044e-01	1.60	0.914877
PC16	5.231928e-01	1.57	0.930607
PC17	5.141585e-01	1.55	0.946065
PC18	4.690880e-01	1.41	0.960169
PC19	3.601734e-01	1.08	0.970997

Komponen Utama	*Eigenvalue*	*Varians Rasio (%)*	*Cumulative Variance Rasio*
PC20	2.272089e-01	0.68	0.977829
PC21	1.899506e-01	0.57	0.983539
PC22	1.849257e-01	0.56	0.989099
PC23	1.505614e-01	0.45	0.993626
PC24	9.850040e-02	0.30	0.996587
PC25	5.767778e-02	0.17	0.998321
PC26	5.117104e-02	0.15	0.999860
PC27	4.658278e-03	0.01	1.000000
PC28	3.463249e-30	0.00	1.000000
PC29	3.733726e-32	0.00	1.000000
PC30	3.733726e-32	0.00	1.000000

Pada tabel 4.5 dapat dilihat bahwa nilai *variance rasio* yang lebih dari 1 dan *cumulative variance rasio*nya lebih dari 95% terdapat bahwa pada PC1 memiliki nilai *variance rasio* tertinggi yaitu 16.87 yang merupakan kontribusi tertinggi terhadap total varians data dan pada *cumulative variance rasio* juga akan meningkat di setiap PCnya. Pada PC1, *cumulative variance ratio* mencapai 0.168743, yang berarti PC1 sendiri dapat menjelaskan sekitar 16.87% dari total varians data. Dan nilai variance rasio diatas 1 terdapat pada PC19 dengan nilai 1.71 dan *cumulative variance ratio* mencapai 0.962717, yang menunjukkan bahwa 19 komponen utama tersebut bersama-sama dapat menjelaskan sekitar 96.62% dari total varians data. Dengan demikian, grafik tersebut menggambarkan bahwa kontirbusi masing-masing PC terhadap total varians data dan meningkatnya kumulatif seiring dengan penambahan PC. Hal ini menunjukkan berapa banyak PC yang akan dipertahankan dalam reduksi dimensi data.

4.3 Hasil Uji Coba

Pada sub-bab ini akan membahas hasil dari uji coba dengan fitur yang sudah direduksi oleh PCA dan diklasifikasikan menggunakan kernel linear pada *Support Vector Machine* (SVM).

4.3.1 Pengujian dengan Metode SVM PCA SMOTE

Pada pengujian ini, akan dilakukan teknik PCA (*Principal Component Analysis*) untuk mereduksi dimensi fitur pada dataset agar dapat mengurasi kompleksitas data dan meningkatkan kinerja model. Dan juga dilakukan teknik SMOTE (*Synthetic Minority Over-sampling Technique*) untuk menangani ketidakseimbangan kelas dalam dataset.

4.3.1.1 Pengujian Model A

Pada pengujian model A, nilai performa dari sistem meliputi *accuracy*, *precision recall,* dan *f1score* yang ditunjukkan pada tabel 4.6 dengan rasio perbandingan data latih dan data uji sebesar 90:10. Pada pengujian ini akan digunakan dalam empat target yaitu *Hinselmann, Schiller, Cytology*, dan *Biopsy*. Dengan pengujian 90:10 ini hasil akurasi paling tinggi didapatkan oleh target *Hinselmann* dengan nilai 75%.

Tabel 4. 6 Performa Sistem Rasio 90:10

Target	Akurasi	Presisi	Recall	f1-Score
Hinselmann	75%	71%	6%	748%
Schiller	62%	70%	44%	54%
Cytology	66%	67%	73%	70%
Biopsy	66%	79%	46%	59%

Hasil perhitungan *accuracy, precision, recall,* dan *f1score* pada target *Hinselmann* yang didapatkan dari evaluasi *confusion matrix* sebagai berikut.

Tabel 4. 7 *Confusion Matrix* Target *Hinselmann* Rasio 90:10

	Data Prediksi	
Data Aktual	**T**	**F**
T	74	12
F	29	50

Pada tabel 4.7 ditunjukkan hasil dari klasifikasi model untuk target *Hinselmann*. Terdapat 74 sampel yang terdeteksi benar positif (*True Positives* - TP), 50 sampel yang terdeteksi benar negatif (*True Negatives* - TN), 29 sampel yang salah terdeteksi positif (*False Positives* - FP), dan 12 sampel positif yang salah terdeteksi negatif (*False Negatives* - FN).Untuk *hyperparamter* pada target *Hinselmann* ini mendapat *hyperparameter* terbaik yaitu *hyperparameter* C 100 dengan akurasi 75%.

Berikut hasil perhitungan dari *accuracy, precision, recall*, dan *f1score* pada target *Schiller* yang didapatkan dari evaluasi *confusion matrix*.

Tabel 4. 8 *Confusion Matrix* Target *Schiller* Rasio 90:10

	Data Prediksi	
Data Aktual	**T**	**F**
T	35	44
F	15	63

Pada tabel 4.8 ditunjukkan hasil dari klasifikasi model untuk target *Schiller*. Terdapat 35 sampel yang terdeteksi benar positif (*True Positives* - TP), 63 sampel yang terdeteksi benar negatif (*True Negatives* - TN), 15 sampel yang salah terdeteksi positif (*False Positives* - FP), dan 44 sampel positif yang salah terdeteksi negatif (*False Negatives* - FN). Untuk *hyperparameter* pada target *Schiller* ini mendapat *hyperparameter* terbaik yaitu *hyperparameter* C 100 dengan akurasi 62%.

Berikut hasil perhitungan dari *accuracy, precision, recall*, dan *f1score* pada target *Cytology* yang didapatkan dari evaluasi *confusion matrix*.

Tabel 4. 9 *Confusion Matrix* Target *Cytology* Rasio 90:10

	Data Prediksi	
Data Aktual	**T**	**F**
T	65	23
F	32	43

Pada tabel 4.5 ditunjukkan hasil dari klasifikasi model untuk target *Citology*. Terdapat 65 sampel yang terdeteksi benar positif (*True Positives* - TP), 43 sampel yang terdeteksi benar negatif (*True Negatives* - TN), 32 sampel yang salah terdeteksi positif (*False Positives* - FP), dan 23 sampel positif yang salah terdeteksi negatif (*False Negatives* - FN). Untuk *hyperparameter* pada target *Cytology* ini mendapat *hyperparameter* terbaik yaitu *hyperparameter* C 10 dengan akurasi 66%.

Berikut hasil perhitungan dari *accuracy, precision, recall*, dan *f1score* pada target *Biopsy* yang didapatkan dari evaluasi *confusion matrix*.

Tabel 4. 10 *Confusion Matrix* Target *Biopsy* Rasio 90:10

	Data Prediksi	
Data Aktual	**T**	**F**
T	39	44
F	10	68

Pada tabel 4.6 ditunjukkan hasil dari klasifikasi model untuk target *Biopsy*. Terdapat 39 sampel yang terdeteksi benar positif (*True Positives* - TP), 68 sampel yang terdeteksi benar negatif (*True Negatives* - TN), 10 sampel yang salah terdeteksi positif (*False Positives* - FP), dan 44 sampel positif yang salah terdeteksi negatif (*False Negatives* - FN). Untuk *hyperparameter* pada target *Biopsy* ini mendapat *hyperparameter* terbaik yaitu *hyperparameter* C 0.1 dengan akurasi 66%.

4.3.1.2 Pengujian Model B

Pada pengujian model B, nilai performa dari sistem meliputi *accuracy*, *precision recall*, dan *f1score* yang ditunjukkan pada tabel 4.11 dengan rasio perbandingan data latih dan data uji sebesar 80:20. Pada pengujian ini akan digunakan dalam empat target yaitu *Hinselmann*, *Schiller*, *Cytology*, dan *Biopsy*. Dengan pengujian 80:20 ini hasil akurasi paling tinggi didapatkan oleh target *Hinselmann* dengan nilai 79%.

Tabel 4. 11 Performa Sistem Rasio 80:20

Target	Akurasi	Presisi	Recall	f1-Score
Hinselmann	79%	76%	89%	82%
Schiller	61%	71%	42%	53%
Cytology	66%	63%	81%	71%
Biopsy	64%	74%	47%	58%

Berikut hasil perhitungan *accuracy, precision, recall*, dan *f1score* pada target *Hinselmann* yang didapatkan dari evaluasi *confusion matrix*.

Tabel 4. 12 *Confusion Matrix* Target *Hinselmann* Rasio 80:20

	Data Prediksi	
Data Aktual	**T**	**F**
T	156	19
F	48	107

Pada tabel 4.12 ditunjukkan hasil dari klasifikasi model untuk target *Hinselmann*. Terdapat 156 sampel yang terdeteksi benar positif (*True Positives* - TP), 107 sampel yang terdeteksi benar negatif (*True Negatives* - TN), 48 sampel yang salah terdeteksi positif (*False Positives* - FP), dan 19 sampel positif yang salah terdeteksi negatif (*False Negatives* - FN). Untuk *hyperparameter* pada target *Hinselmann* ini mendapat *hyperparameter* terbaik yaitu *hyperparameter* C 100 dengan akurasi 79%.

Berikut hasil perhitungan dari *accuracy, precision, recall*, dan *f1score* pada target *Schiller* yang didapatkan dari evaluasi *confusion matrix*.

Tabel 4. 13 *Confusion Matrix* Target *Schiller* Rasio 80:20

	Data Prediksi	
Data Aktual	**T**	**F**
T	69	93
F	28	124

Pada tabel 4.13 ditunjukkan hasil dari klasifikasi model untuk target *Schiller*. Terdapat 69 sampel yang terdeteksi benar positif (*True Positives* - TP), 124 sampel yang terdeteksi benar negatif (*True Negatives* - TN), 28 sampel yang salah terdeteksi positif (*False Positives* - FP), dan 93 sampel positif yang salah terdeteksi negatif (*False Negatives* - FN). Untuk *hyperparameter* pada target *Schiller* ini mendapat *hyperparameter* terbaik yaitu *hyperparameter* C 100 dengan akurasi 61%.

Berikut hasil perhitungan dari *accuracy, precision, recall*, dan *f1score* pada target *Cytology* yang didapatkan dari evaluasi *confusion matrix*.

Tabel 4. 14 *Confusion Matrix* Target *Cytology* Rasio 80:20

	Data Prediksi	
Data Aktual	**T**	**F**
T	135	30
F	78	83

Pada tabel 4.14 ditunjukkan hasil dari klasifikasi model untuk target *Citology.* Terdapat 135 sampel yang terdeteksi benar positif (*True Positives* - TP), 83 sampel yang terdeteksi benar negatif (*True Negatives* - TN), 78 sampel yang salah terdeteksi positif (*False Positives* - FP), dan 30 sampel positif yang salah terdeteksi negatif (*False Negatives* - FN). Untuk *hyperparameter* pada target

Citology ini mendapat *hyperparameter* terbaik yaitu *hyperparameter* C 100 dengan akurasi 66%.

Berikut hasil perhitungan dari *accuracy, precision, recall*, dan *f1score* pada target *Biopsy* yang didapatkan dari evaluasi *confusion matrix*.

Tabel 4. 15 *Confusion Matrix* Target *Biopsy* Rasio 80:20

	Data Prediksi	
Data Aktual	**T**	**F**
T	79	86
F	27	130

Pada tabel 4.15 ditunjukkan hasil dari klasifikasi model untuk target *Biopsy*. Terdapat 79 sampel yang terdeteksi benar positif (*True Positves* - TP), 130 sampel yang terdeteksi benar negatif (*True Negatives* - TN), 27 sampel yang salah terdeteksi positif (*False Positives* - FP), dan 86 sampel positif yang salah terdeteksi negatif (*False Negatives* - FN). Untuk *hyperparameter* pada target *Biopsy* ini mendapat *hyperparameter* terbaik yaitu *hyperparameter* C 10 dengan akurasi 64%.

4.3.1.3 Pengujian Model C

Pada pengujian model C, nilai performa dari sistem meliputi *accuracy, precision recall,* dan *f1score* yang ditunjukkan pada tabel 4.16 dengan rasio perbandingan data latih dan data uji sebesar 70:30. Pada pengujian ini akan digunakan dalam empat target yaitu *Hinselmann, Schiller, Cytology*, dan *Biopsy*. Dengan pengujian 70:30 ini hasil akurasi paling tinggi didapatkan oleh target *Hinselmann* dengan nilai 76%.

Tabel 4. 16 Performa Sistem Rasio 70:30

Target	**Akurasi**	**Presisi**	**Recall**	**f1-Score**
Hinselmann	76%	71%	90%	79%
Schiller	61%	72%	43%	54%
Cytology	65%	63%	78%	70%
Biopsy	65%	75%	47%	58%

Hasil perhitungan *accuracy, precision, recall*, dan *f1score* pada target *Hinselmann* yang didapatkan dari evaluasi *confusion matrix* sebagai berikut.

Tabel 4. 17 *Confusion Matrix* Target *Hinselmann* Rasio 70:30

	Data Prediksi	
Data Aktual	**T**	**F**
T	225	24
F	91	154

Pada tabel 4.17 ditunjukkan hasil dari klasifikasi model untuk target *Hinselmann*. Terdapat 225 sampel yang terdeteksi benar positif (*True Positves* - TP), 154 sampel yang terdeteksi benar negatif (*True Negatives* - TN), 91 sampel yang salah terdeteksi positif (*False Positives* - FP), dan 24 sampel positif yang salah terdeteksi negatif (*False Negatives* - FN). Untuk *hyperparameter* pada target *Hinselmann* ini mendapat *hyperparameter* terbaik yaitu *hyperparameter* C 100 dengan akurasi 76%.

Berikut hasil perhitungan dari *accuracy, precision, recall*, dan *f1score* pada target *Schiller* yang didapatkan dari evaluasi *confusion matrix*.

Tabel 4. 18 *Confusion Matrix* Target *Schiller* Rasio 70:30

	Data Prediksi	
Data Aktual	**T**	**F**
T	106	135
F	40	190

Pada tabel 4.18 ditunjukkan hasil dari klasifikasi model untuk target *Schiller*. Terdapat 106 sampel yang terdeteksi benar positif (*True Positves* - TP), 190 sampel yang terdeteksi benar negatif (*True Negatives* - TN), 40 sampel yang salah terdeteksi positif (*False Positives* - FP), dan 135 sampel positif yang salah terdeteksi negatif (*False Negatives* - FN). Untuk *hyperparameter* pada target

Schiller ini mendapat *hyperparameter* terbaik yaitu *hyperparameter* C 100 dengan akurasi 61%.

Berikut hasil perhitungan dari *accuracy, precision, recall*, dan *f1score* pada target *Cytology* yang didapatkan dari evaluasi *confusion matrix*.

Tabel 4. 19 *Confusion Matrix* Target *Cytology* Rasio 70:30

	Data Prediksi	
Data Aktual	**T**	**F**
T	196	54
F	114	125

Pada tabel 4.19 ditunjukkan hasil dari klasifikasi model untuk target *Citology*. Terdapat 196 sampel yang terdeteksi benar positif (*True Positves* - TP), 125 sampel yang terdeteksi benar negatif (*True Negatives* - TN), 114 sampel yang salah terdeteksi positif (*False Positives* - FP), dan 54 sampel positif yang salah terdeteksi negatif (*False Negatives* - FN). Untuk *hyperparameter* pada target *Citology* ini mendapat *hyperparameter* terbaik yaitu *hyperparameter* C 100 dengan akurasi 65%.

Berikut hasil perhitungan dari *accuracy, precision, recall*, dan *f1score* pada target *Biopsy* yang didapatkan dari evaluasi *confusion matrix*.

Tabel 4. 20 *Confusion Matrix* Target *Biopsy* Rasio 70:30

	Data Prediksi	
Data Aktual	**T**	**F**
T	117	128
F	39	198

Pada tabel 4.20 ditunjukkan hasil dari klasifikasi model untuk target *Biopsy*. Terdapat 117 sampel yang terdeteksi benar positif (*True Positves* - TP), 198 sampel yang terdeteksi benar negatif (*True Negatives* - TN), 39 sampel yang salah terdeteksi positif (*False Positives* - FP), dan 128 sampel positif yang salah

terdeteksi negatif (*False Negatives* - FN). Untuk *hyperparameter* pada target *Biopsy* ini mendapat *hyperparameter* terbaik yaitu *hyperparameter* C 10 dengan akurasi 65%.

4.3.1.4 Pengujian Model D

Pada pengujian model D, nilai performa dari sistem meliputi *accuracy, precision recall,* dan *f1score* yang ditunjukkan pada tabel 4.21 dengan rasio perbandingan data latih dan data uji sebesar 60:40. Pada pengujian ini akan digunakan dalam empat target yaitu *Hinselmann, Schiller, Cytology*, dan *Biopsy*. Dengan pengujian 60:40 ini hasil akurasi paling tinggi didapatkan oleh target *Hinselmann* dengan nilai 74%.

Tabel 4. 21 Performa Sistem Rasio 60:40

Target	Akurasi	Presisi	Recall	f1-Score
Hinselmann	74%	69%	89%	77%
Schiller	63%	73%	42%	54%
Cytology	62%	59%	75%	66%
Biopsy	47%	75%	47%	58%

Hasil perhitungan *accuracy*, *precision, recall*, dan *f1score* pada target *Hinselmann* yang didapatkan dari evaluasi *confusion matrix* sebagai berikut.

Tabel 4. 22 *Confusion Matrix* Target *Hinselmann* Rasio 60:40

	Data Prediksi	
Data Aktual	**T**	**F**
T	292	35
F	130	202

Pada tabel 4.22 ditunjukkan hasil dari klasifikasi model untuk target *Hinselmann*. Terdapat 292 sampel yang terdeteksi benar positif (*True Positves* - TP), 202 sampel yang terdeteksi benar negatif (*True Negatives* - TN), 130 sampel yang salah terdeteksi positif (*False Positives* - FP), dan 35 sampel positif yang salah

terdeteksi negatif (*False Negatives* - FN). Untuk *hyperparameter* pada target *Hinselmann* ini mendapat *hyperparameter* terbaik yaitu *hyperparameter* C 10 dengan akurasi 74%.

Berikut hasil perhitungan dari *accuracy, precision, recall*, dan *f1score* pada target *Schiller* yang didapatkan dari evaluasi *confusion matrix*.

Tabel 4. 23 Confusion Matrix Target Schiller Rasio 60:40

	Data Prediksi	
Data Aktual	**T**	**F**
T	134	179
F	49	266

Pada tabel 4.23 ditunjukkan hasil dari klasifikasi model untuk target *Schiller*. Terdapat 134 sampel yang terdeteksi benar positif (*True Positves* - TP), 266 sampel yang terdeteksi benar negatif (*True Negatives* - TN), 49 sampel yang salah terdeteksi positif (*False Positives* - FP), dan 179 sampel positif yang salah terdeteksi negatif (*False Negatives* - FN). Untuk *hyperparameter* pada target *Schiller* ini mendapat *hyperparameter* terbaik yaitu *hyperparameter* C 100 dengan akurasi 63%.

Berikut hasil perhitungan dari *accuracy, precision, recall*, dan *f1score* pada target *Cytology* yang didapatkan dari evaluasi *confusion matrix*.

Tabel 4. 24 *Confusion Matrix* Target *Cytology* Rasio 60:40

	Data Prediksi	
Data Aktual	**T**	**F**
T	246	79
F	167	160

Pada tabel 4.24 ditunjukkan hasil dari klasifikasi model untuk target *Citology*. Terdapat 246 sampel yang terdeteksi benar positif (*True Positves* - TP), 160 sampel yang terdeteksi benar negatif (*True Negatives* - TN), 167 sampel yang

salah terdeteksi positif (*False Positives* - FP), dan 79 sampel positif yang salah terdeteksi negatif (*False Negatives* - FN). Untuk *hyperparameter* pada target *Citology* ini mendapat *hyperparameter* terbaik yaitu *hyperparameter* C 10 dengan akurasi 62%.

Berikut hasil perhitungan dari *accuracy, precision, recall,* dan *f1score* pada target *Biopsy* yang didapatkan dari evaluasi *confusion matrix*.

Tabel 4. 25 *Confusion Matrix* Target *Biopsy* Rasio 60:40

	Data Prediksi	
Data Aktual	**T**	**F**
T	155	168
F	50	270

Pada tabel 4.25 ditunjukkan hasil dari klasifikasi model untuk target *Biopsy*. Terdapat 155 sampel yang terdeteksi benar positif (*True Positves* - TP), 270 sampel yang terdeteksi benar negatif (*True Negatives* - TN), 50 sampel yang salah terdeteksi positif (*False Positives* - FP), dan 168 sampel positif yang salah terdeteksi negatif (*False Negatives* - FN). Untuk *hyperparameter* pada target *Biopsy* ini mendapat *hyperparameter* terbaik yaitu *hyperparameter* C 10 dengan akurasi 66%.

4.3.2 Pengujian dengan Metode SVM SMOTE tanpa PCA

Pada pengujian ini, dalam melakukan klasifikasi kanker serviks menggunakan metode SVM tanpa menggunakan teknik PCA untuk mereduksi dimensi fitur. Akan digunakan seluruh fitur yang ada dalam dataset untuk model klasifikasi dan juga dilakukan teknik SMOTE (*Synthetic Minority Over-sampling Technique*) untuk menangani ketidakseimbangan kelas dalam dataset.

4.3.2.1 Pengujian Model A

Pada pengujian model A, nilai performa dari sistem meliputi *accuracy*, *precision recall*, dan *f1score* yang ditunjukkan pada tabel 4.26 dengan rasio perbandingan data latih dan data uji sebesar 90:10. Pada pengujian ini akan digunakan dalam empat target yaitu *Hinselmann*, *Schiller*, *Cytology*, dan *Biopsy*. Dengan pengujian 90:10 ini hasil akurasi paling tinggi didapatkan oleh target *Schiller* dengan nilai 75%.

Tabel 4. 26 Performa Sistem Rasio 90:10 tanpa PCA

Target	Akurasi	Presisi	Recall	f1-Score
Hinselmann	61%	71%	42%	53%
Schiller	75%	71%	86%	7%
Cytology	69%	67%	73%	70%
Biopsy	66%	79%	46%	59%

Hasil perhitungan *accuracy, precision, recall*, dan *f1score* pada target *Hinselmann* yang didapatkan dari evaluasi confusion matrix sebagai berikut.

Tabel 4. 27 *Confusion Matrix* Target *Hinselmann* Rasio 90:10 tanpa PCA

	Data Prediksi	
Data Aktual	**T**	**F**
T	69	93
F	28	24

Pada tabel 4.27 ditunjukkan hasil dari klasifikasi model untuk target *Hinselmann.* Terdapat 69 sampel yang terdeteksi benar positif (*True Positves* - TP), 24 sampel yang terdeteksi benar negatif (*True Negatives* - TN), 28 sampel yang salah terdeteksi positif (*False Positives* - FP), dan 93 sampel positif yang salah terdeteksi negatif (*False Negatives* - FN). Untuk *hyperparameter* pada target *Hinselmann* ini mendapat *hyperparameter* terbaik yaitu *hyperparameter* C 100 dengan akurasi 61%.

Berikut hasil perhitungan dari *accuracy, precision, recall,* dan *f1score* pada target *Schiller* yang didapatkan dari evaluasi confusion matrix.

Tabel 4. 28 *Confusion Matrix* Target *Schiller* Rasio 90:10 tanpa PCA

	Data Prediksi	
Data Aktual	**T**	**F**
T	74	12
F	29	50

Pada tabel 4.28 ditunjukkan hasil dari klasifikasi model untuk target *Schiller*. Terdapat 74 sampel yang terdeteksi benar positif (*True Positves* - TP), 50 sampel yang terdeteksi benar negatif (*True Negatives* - TN), 29sampel yang salah terdeteksi positif (*False Positives* - FP), dan 12 sampel positif yang salah terdeteksi negatif (*False Negatives* - FN). Untuk *hyperparameter* pada target *Schiller* ini mendapat *hyperparameter* terbaik yaitu *hyperparameter* C 100 dengan akurasi 75%.

Berikut hasil perhitungan dari *accuracy, precision, recall,* dan *f1score* pada target *Cytology* yang didapatkan dari evaluasi *confusion matrix*.

Tabel 4. 29 *Confusion Matrix* Target *Cytology* Rasio 90:10 tanpa PCA

	Data Prediksi	
Data Aktual	**T**	**F**
T	65	23
F	32	43

Pada tabel 4.29 ditunjukkan hasil dari klasifikasi model untuk target *Citology*. Terdapat 65 sampel yang terdeteksi benar positif (*True Positves* - TP), 43 sampel yang terdeteksi benar negatif (*True Negatives* - TN), 32 sampel yang salah terdeteksi positif (*False Positives* - FP), dan 23 sampel positif yang salah terdeteksi negatif (*False Negatives* - FN). Untuk *hyperparameter* pada target *Citology* ini mendapat *hyperparameter* terbaik yaitu *hyperparameter* C 10 dengan akurasi 66%.

Berikut hasil perhitungan dari *accuracy, precision, recall*, dan *f1score* pada target *Biopsy* yang didapatkan dari evaluasi *confusion matrix*.

Tabel 4. 30 *Confusion Matrix* Target *Biopsy* Rasio 90:10

	Data Prediksi	
Data Aktual	**T**	**F**
T	39	44
F	10	68

Pada tabel 4.30 ditunjukkan hasil dari klasifikasi model untuk target *Biopsy*. Terdapat 39 sampel yang terdeteksi benar positif (*True Positves* - TP), 68 sampel yang terdeteksi benar negatif (*True Negatives* - TN), 10 sampel yang salah terdeteksi positif (*False Positives* - FP), dan 44 sampel positif yang salah terdeteksi negatif (*False Negatives* - FN). Untuk *hyperparameter* pada target *Biopsy* ini mendapat *hyperparameter* terbaik yaitu *hyperparameter* C 0.1 dengan akurasi 66%.

4.3.2.2 Pengujian Model B

Pada pengujian model B, nilai performa dari sistem meliputi *accuracy, precision recall*, dan *f1score* yang ditunjukkan pada tabel 4.31 dengan rasio perbandingan data latih dan data uji sebesar 80:20. Pada pengujian ini akan digunakan dalam empat target yaitu *Hinselmann, Schiller, Cytology*, dan *Biopsy*. Dengan pengujian 80:20 ini hasil akurasi paling tinggi didapatkan oleh target *Hinselmann* dengan nilai 79%.

Tabel 4. 31 Performa Sistem Rasio 80:20

Target	**Akurasi**	**Presisi**	**Recall**	**f1-Score**
Hinselmann	79%	76%	89%	82%
Schiller	62%	71%	42%	53%
Cytology	66%	63%	81%	71%
Biopsy	65%	76%	46%	57%

Hasil perhitungan *accuracy*, *precision*, *recall*, dan *f1score* pada target *Hinselmann* yang didapatkan dari evaluasi *confusion matrix* sebagai berikut.

Tabel 4. 32 *Confusion Matrix* Target *Hinselmann* Rasio 80:20 tanpa PCA

	Data Prediksi	
Data Aktual	**T**	**F**
T	156	19
F	48	107

Pada tabel 4.32 ditunjukkan hasil dari klasifikasi model untuk target *Hinselmann*. Terdapat 156 sampel yang terdeteksi benar positif (*True Positves* - TP), 107 sampel yang terdeteksi benar negatif (*True Negatives* - TN), 48 sampel yang salah terdeteksi positif (*False Positives* - FP), dan 19 sampel positif yang salah terdeteksi negatif (*False Negatives* - FN). Untuk *hyperparameter* pada target *Hinselmann* ini mendapat *hyperparameter* terbaik yaitu *hyperparameter* C 100 dengan akurasi 79%.

Berikut hasil perhitungan dari *accuracy*, *precision*, *recall*, dan *f1score* pada target *Schiller* yang didapatkan dari evaluasi *confusion matrix*.

Tabel 4. 33 *Confusion Matrix* Target *Schiller* Rasio 80:20 tanpa PCA

	Data Prediksi	
Data Aktual	**T**	**F**
T	69	93
F	28	124

Pada tabel 4.33 ditunjukkan hasil dari klasifikasi model untuk target *Schiller*. Terdapat 69 sampel yang terdeteksi benar positif (*True Positves* - TP), 124 sampel yang terdeteksi benar negatif (*True Negatives* - TN), 28 sampel yang salah terdeteksi positif (*False Positives* - FP), dan 93 sampel positif yang salah terdeteksi negatif (*False Negatives* - FN). Untuk *hyperparameter* pada target *Schiller* ini

mendapat *hyperparameter* terbaik yaitu *hyperparameter* C 100 dengan akurasi 61%.

Berikut hasil perhitungan dari *accuracy, precision, recall*, dan *f1score* pada target *Cytology* yang didapatkan dari evaluasi *confusion matrix*.

Tabel 4. 34 *Confusion Matrix* Target *Cytology* Rasio 80:20 tanpa PCA

	Data Prediksi	
Data Aktual	**T**	**F**
T	135	30
F	78	83

Pada tabel 4.33 ditunjukkan hasil dari klasifikasi model untuk target *Citology*. Terdapat 135 sampel yang terdeteksi benar positif (*True Positves* - TP), 83 sampel yang terdeteksi benar negatif (*True Negatives* - TN), 78 sampel yang salah terdeteksi positif (*False Positives* - FP), dan 30 sampel positif yang salah terdeteksi negatif (*False Negatives* - FN). Untuk *hyperparameter* pada target *Citology* ini mendapat *hyperparameter* terbaik yaitu *hyperparameter* C 100 dengan akurasi 66%.

Berikut hasil perhitungan dari *accuracy, precision, recall*, dan *f1score* pada target *Biopsy* yang didapatkan dari evaluasi *confusion matrix*.

Tabel 4. 35 *Confusion Matrix* Target *Biopsy* Rasio 80:20 tanpa PCA

	Data Prediksi	
Data Aktual	**T**	**F**
T	77	88
F	24	133

Pada tabel 4.35 ditunjukkan hasil dari klasifikasi model untuk target *Biopsy*. Terdapat 77 sampel yang terdeteksi benar positif (*True Positves* - TP), 133 sampel yang terdeteksi benar negatif (*True Negatives* - TN), 24 sampel yang salah terdeteksi positif (*False Positives* - FP), dan 88 sampel positif yang salah terdeteksi

negatif (*False Negatives* - FN). Untuk *hyperparameter* pada target *Biopsy* ini mendapat *hyperparameter* terbaik yaitu *hyperparameter* C 0.1 dengan akurasi 65%.

4.3.2.3 Pengujian Model C

Pada pengujian model C, nilai performa dari sistem meliputi *accuracy, precision recall,* dan *f1score* yang ditunjukkan pada tabel 4.32 dengan rasio perbandingan data latih dan data uji sebesar 70:30. Pada pengujian ini akan digunakan dalam 4 target yaitu *Hinselmann, Schiller, Cytology*, dan *Biopsy*. Dengan pengujian 70:30 ini hasil akurasi paling tinggi didapatkan oleh target *Hinselmann* dengan nilai 76%.

Tabel 4. 36 Performa Sistem Rasio 70:30 tanpa PCA

Target	Akurasi	Presisi	Recall	f1-Score
Hinselmann	76%	71%	90%	79%
Schiller	62%	72%	43%	54%
Cytology	65%	63%	78%	69%
Biopsy	65%	75%	47%	58%

Hasil perhitungan *accuracy, precision, recall*, dan *f1score* pada target *Hinselmann* yang didapatkan dari evaluasi *confusion matrix* sebagai berikut.

Tabel 4. 37 *Confusion Matrix* Target *Hinselmann* Rasio 70:30 tanpa PCA

	Data Prediksi	
Data Aktual	**T**	**F**
T	225	24
F	91	154

Pada tabel 4.37 ditunjukkan hasil dari klasifikasi model untuk target *Hinselmann*. Terdapat 225 sampel yang terdeteksi benar positif (*True Positves* - TP), 154 sampel yang terdeteksi benar negatif (*True Negatives* - TN), 91 sampel yang salah terdeteksi positif (*False Positives* - FP), dan 24 sampel positif yang salah terdeteksi negatif (*False Negatives* - FN). Untuk *hyperparameter* pada target

Hinselmann ini mendapat *hyperparameter* terbaik yaitu *hyperparameter* C 100 dengan akurasi 76%.

Berikut hasil perhitungan dari *accuracy, precision, recall*, dan *f1score* pada target *Schiller* yang didapatkan dari evaluasi *confusion matrix*.

Tabel 4. 38 *Confusion Matrix* Target *Schiller* Rasio 70:30 tanpa PCA

	Data Prediksi	
Data Aktual	**T**	**F**
T	106	135
F	40	190

Pada tabel 4.38 ditunjukkan hasil dari klasifikasi model untuk target *Schiller*. Terdapat 106 sampel yang terdeteksi benar positif (*True Positves* - TP), 190 sampel yang terdeteksi benar negatif (*True Negatives* - TN), 40 sampel yang salah terdeteksi positif (*False Positives* - FP), dan 135 sampel positif yang salah terdeteksi negatif (*False Negatives* - FN). Untuk *hyperparameter* pada target *Schiller* ini mendapat *hyperparameter* terbaik yaitu *hyperparameter* C 100 dengan akurasi 62%.

Berikut hasil perhitungan dari *accuracy, precision, recall*, dan *f1score* pada target *Cytology* yang didapatkan dari evaluasi *confusion matrix*.

Tabel 4. 39 *Confusion Matrix* Target *Cytology* Rasio 70:30 tanpa PCA

	Data Prediksi	
Data Aktual	**T**	**F**
T	195	55
F	113	126

Pada tabel 4.39 ditunjukkan hasil dari klasifikasi model untuk target *Citology*. Terdapat 195 sampel yang terdeteksi benar positif (*True Positves* - TP), 126 sampel yang terdeteksi benar negatif (*True Negatives* - TN), 113 sampel yang salah terdeteksi positif (*False Positives* - FP), dan 55 sampel positif yang salah

terdeteksi negatif (*False Negatives* - FN). Untuk *hyperparameter* pada target *Citology* ini mendapat *hyperparameter* terbaik yaitu *hyperparameter* C 100 dengan akurasi 65%.

Berikut hasil perhitungan dari *accuracy, precision, recall*, dan *f1score* pada target *Biopsy* yang didapatkan dari evaluasi *confusion matrix*.

Tabel 4. 40 *Confusion Matrix* Target *Biopsy* Rasio 70:30 tanpa PCA

	Data Prediksi	
Data Aktual	**T**	**F**
T	117	128
F	39	198

Pada tabel 4.40 ditunjukkan hasil dari klasifikasi model untuk target *Biopsy*. Terdapat 117 sampel yang terdeteksi benar positif (*True Positves* - TP), 198 sampel yang terdeteksi benar negatif (*True Negatives* - TN), 39 sampel yang salah terdeteksi positif (*False Positives* - FP), dan 128 sampel positif yang salah terdeteksi negatif (*False Negatives* - FN). Untuk *hyperparameter* pada target *Biopsy* ini mendapat *hyperparameter* terbaik yaitu *hyperparameter* C 10 dengan akurasi 65%.

4.3.2.4 Pengujian Model D

Pada pengujian model D, nilai performa dari sistem meliputi *accuracy, precision recall*, dan *f1score* yang ditunjukkan pada tabel 4.41 dengan rasio perbandingan data latih dan data uji sebesar 60:40. Pada pengujian ini akan digunakan dalam empat target yaitu *Hinselmann, Schiller, Cytology*, dan *Biopsy*. Dengan pengujian 60:40 ini hasil akurasi paling tinggi didapatkan oleh target *Hinselmann* dengan nilai 74%.

Tabel 4. 41 Performa Sistem Rasio 60:40 tanpa PCA

Target	Akurasi	Presisi	Recall	f1-Score
Hinselmann	74%	69%	89%	77%
Schiller	63%	73%	41%	54%
Cytology	62%	59%	75%	66%
Biopsy	66%	75%	47%	58%

Hasil perhitungan *accuracy, precision, recall*, dan *f1score* pada target *Hinselmann* yang didapatkan dari evaluasi *confusion matrix* sebagai berikut.

Tabel 4. 42 *Confusion Matrix* Target *Hinselmann* Rasio 60:40 tanpa PCA

	Data Prediksi	
Data Aktual	**T**	**F**
T	292	35
F	130	102

Pada tabel 4.42 ditunjukkan hasil dari klasifikasi model untuk target *Hinselmann*. Terdapat 292 sampel yang terdeteksi benar positif (*True Positves* - TP), 102 sampel yang terdeteksi benar negatif (*True Negatives* - TN), 130 sampel yang salah terdeteksi positif (*False Positives* - FP), dan 35 sampel positif yang salah terdeteksi negatif (*False Negatives* - FN). Untuk *hyperparameter* pada target *Hinselmann* ini mendapat *hyperparameter* terbaik yaitu *hyperparameter* C 10 dengan akurasi 74%. Hasil perhitungan dari *accuracy, precision, recall*, dan *f1score* pada target *Schiller* yang didapatkan dari evaluasi *confusion matrix*.

Tabel 4. 43 *Confusion Matrix* Target *Schiller* Rasio 60:40 tanpa PCA

	Data Prediksi	
Data Aktual	**T**	**F**
T	134	179
F	48	267

Pada tabel 4.43 ditunjukkan hasil dari klasifikasi model untuk target *Schiller*. Terdapat 134 sampel yang terdeteksi benar positif (*True Positves* - TP), 267sampel yang terdeteksi benar negatif (*True Negatives* - TN), 48 sampel yang

salah terdeteksi positif (*False Positives* - FP), dan 179 sampel positif yang salah terdeteksi negatif (*False Negatives* - FN). Untuk *hyperparameter* pada target *Schiller* ini mendapat *hyperparameter* terbaik yaitu *hyperparameter* C 10 dengan akurasi 63%.

Berikut hasil perhitungan dari *accuracy, precision, recall,* dan *f1score* pada target *Cytology* yang didapatkan dari evaluasi *confusion matrix*.

Tabel 4. 44 *Confusion Matrix* Target *Cytology* Rasio 60:40 tanpa PCA

	Data Prediksi	
Data Aktual	**T**	**F**
T	246	79
F	167	160

Pada tabel 4.44 ditunjukkan hasil dari klasifikasi model untuk target *Citology*. Terdapat 246 sampel yang terdeteksi benar positif (*True Positves* - TP), 160 sampel yang terdeteksi benar negatif (*True Negatives* - TN), 167 sampel yang salah terdeteksi positif (*False Positives* - FP), dan 79 sampel positif yang salah terdeteksi negatif (*False Negatives* - FN). Untuk *hyperparameter* pada target *Citology* ini mendapat *hyperparameter* terbaik yaitu *hyperparameter* C 10 dengan akurasi 62%.

Berikut hasil perhitungan dari *accuracy, precision, recall,* dan *f1score* pada target B*iopsy* yang didapatkan dari evaluasi *confusion matrix*.

Tabel 4. 45 *Confusion Matrix* Target *Biopsy* Rasio 60:40 tanpa PCA

	Data Prediksi	
Data Aktual	**T**	**F**
T	155	168
F	50	270

Pada tabel 4.45 ditunjukkan hasil dari klasifikasi model untuk target *Biopsy*. Terdapat 155 sampel yang terdeteksi benar positif (*True Positves* - TP), 270 sampel

yang terdeteksi benar negatif (*True Negatives* - TN), 50 sampel yang salah terdeteksi positif (*False Positives* - FP), dan 168 sampel positif yang salah terdeteksi negatif (*False Negatives* - FN). Untuk *hyperparameter* pada target *Biopsy* ini mendapat *hyperparameter* terbaik yaitu *hyperparameter* C 10 dengan akurasi 66%.

4.3.3 Pengujian dengan Metode SVM PCA tanpa SMOTE

Pada pengujian ini, akan dilakukan teknik PCA (*Principal Component Analysis*) untuk mereduksi dimensi fitur pada dataset agar dapat mengurasi kompleksitas data dan meningkatkan kinerja model.

4.3.3.1 Pengujian Model A

Pada pengujian model A, nilai performa dari sistem meliputi *accuracy, precision recall,* dan *f1score* yang ditunjukkan pada tabel 4.46 dengan rasio perbandingan data latih dan data uji sebesar 90:10. Pada pengujian ini akan digunakan dalam empat target yaitu *Hinselmann, Schiller, Cytology*, dan *Biopsy*. Dengan pengujian 90:10 ini hasil akurasi paling tinggi didapatkan oleh target *Citology* dengan nilai 98%.

Tabel 4. 46 Performa Sistem Rasio 90:10

Target	Akurasi	Presisi	Recall	f1-Score
Hinselmann	97%	0%	0%	0%
Schiller	94%	0%	0%	0%
Cytology	98%	0%	0%	0%
Biopsy	97%	0%	0%	0%

Hasil perhitungan *accuracy, precision, recall,* dan *f1score* pada target *Hinselmann* yang didapatkan dari evaluasi *confusion matrix* sebagai berikut.

Tabel 4. 47 *Confusion Matrix* Target *Hinselmann* Rasio 90:10 dengan Metode SVM PCA tanpa SMOTE

	Data Prediksi	
Data Aktual	**T**	**F**
T	0	2
F	0	84

Pada tabel 4.47 ditunjukkan hasil dari klasifikasi model untuk target *Hinselmann*. Tidak ada sampel yang terdeteksi benar positif (*True Positives* - TP), 84 sampel yang terdeteksi benar negatif (*True Negatives* - TN), tidak ada sampel yang salah terdeteksi positif (*False Positives* - FP), dan 2 sampel positif yang salah terdeteksi negatif (*False Negatives* - FN). Untuk *hyperparamter* pada target *Hinselmann* ini mendapat *hyperparameter* terbaik yaitu *hyperparameter* C 0.1 dengan akurasi 96%.

Berikut hasil perhitungan dari *accuracy*, *precision, recall*, dan *f1score* pada target *Schiller* yang didapatkan dari evaluasi *confusion matrix*.

Tabel 4. 48 *Confusion Matrix* Target *Schiller* Rasio 90:10

	Data Prediksi	
Data Aktual	**T**	**F**
T	0	5
F	0	81

Pada tabel 4.48 ditunjukkan hasil dari klasifikasi model untuk target *Schiller*. Tidak ada sampel yang terdeteksi benar positif (*True Positives* - TP), 81 sampel yang terdeteksi benar negatif (*True Negatives* - TN), tidak ada sampel yang salah terdeteksi positif (*False Positives* - FP), dan 5 sampel positif yang salah terdeteksi negatif (*False Negatives* - FN). Untuk *hyperparamter* pada target *Schiller* ini mendapat *hyperparameter* terbaik yaitu *hyperparameter* C 0.1 dengan akurasi 94%.

Berikut hasil perhitungan dari *accuracy, precision, recall*, dan *f1score* pada target *Cytology* yang didapatkan dari evaluasi *confusion matrix*.

Tabel 4. 49 *Confusion Matrix* Target *Cytology* Rasio 90:10

	Data Prediksi	
Data Aktual	**T**	**F**
T	0	1
F	0	85

Pada tabel 4.49 ditunjukkan hasil dari klasifikasi model untuk target *Citology*. Tidak ada sampel yang terdeteksi benar positif (*True Positives* - TP), 85 sampel yang terdeteksi benar negatif (*True Negatives* - TN), tidak ada sampel yang salah terdeteksi positif (*False Positives* - FP), dan 1 sampel positif yang salah terdeteksi negatif (*False Negatives* - FN). Untuk *hyperparamter* pada target *Citology* ini mendapat *hyperparameter* terbaik yaitu *hyperparameter* C 0.1 dengan akurasi 98%.

Berikut hasil perhitungan dari *accuracy, precision, recall*, dan *f1score* pada target *Biopsy* yang didapatkan dari evaluasi *confusion matrix*.

Tabel 4. 50 *Confusion Matrix* Target *Biopsy* Rasio 90:10

	Data Prediksi	
Data Aktual	**T**	**F**
T	0	2
F	0	84

Pada tabel 4.50 ditunjukkan hasil dari klasifikasi model untuk target *Biopsy*. Tidak ada sampel yang terdeteksi benar positif (*True Positives* - TP), 84 sampel yang terdeteksi benar negatif (*True Negatives* - TN), tidak ada sampel yang salah terdeteksi positif (*False Positives* - FP), dan 2 sampel positif yang salah terdeteksi negatif (*False Negatives* - FN). Untuk *hyperparamter* pada target *Biopsy* ini mendapat *hyperparameter* terbaik yaitu *hyperparameter* C 0.1 dengan akurasi 97%.

4.3.3.2 Pengujian Model B

Pada pengujian model B, nilai performa dari sistem meliputi *accuracy*, *precision recall*, dan *f1score* yang ditunjukkan pada tabel 4.51 dengan rasio perbandingan data latih dan data uji sebesar 80:20. Pada pengujian ini akan digunakan dalam empat target yaitu *Hinselmann*, *Schiller*, *Cytology*, dan *Biopsy*. Dengan pengujian 80:20 ini hasil akurasi paling tinggi didapatkan oleh target *Citology* dengan nilai 78%.

Tabel 4. 51 Performa Sistem Rasio 80:20

Target	Akurasi	Presisi	Recall	f1-Score
Hinselmann	95%	0%	0%	0%
Schiller	90%	0%	0%	0%
Cytology	98%	0%	0%	0%
Biopsy	94%	100%	6%	16%

Berikut hasil perhitungan *accuracy, precision, recall*, dan *f1score* pada target *Hinselmann* yang didapatkan dari evaluasi *confusion matrix*.

Tabel 4. 52 *Confusion Matrix* Target *Hinselmann* Rasio 80:20

	Data Prediksi	
Data Aktual	**T**	**F**
T	0	8
F	0	164

Pada tabel 4.52 ditunjukkan hasil dari klasifikasi model untuk target *Hinselmann*. Tidak ada sampel yang terdeteksi benar positif (*True Positives* - TP), 164 sampel yang terdeteksi benar negatif (*True Negatives* - TN), tidak ada sampel yang salah terdeteksi positif (*False Positives* - FP), dan 8 sampel positif yang salah terdeteksi negatif (*False Negatives* - FN). Untuk *hyperparamter* pada target *Hinselmann* ini mendapat *hyperparameter* terbaik yaitu *hyperparameter* C 0.1 dengan akurasi 95%.

Berikut hasil perhitungan dari *accuracy, precision, recall*, dan *f1score* pada target *Schiller* yang didapatkan dari evaluasi *confusion matrix*.

Tabel 4. 53 *Confusion Matrix* Target *Schiller* Rasio 80:20

	Data Prediksi	
Data Aktual	**T**	**F**
T	0	17
F	0	155

Pada tabel 4.53 ditunjukkan hasil dari klasifikasi model untuk target *Schiller*. Tidak ada sampel yang terdeteksi benar positif (*True Positives* - TP), 155 sampel yang terdeteksi benar negatif (*True Negatives* - TN), tidak ada sampel yang salah terdeteksi positif (*False Positives* - FP), dan 17 sampel positif yang salah terdeteksi negatif (*False Negatives* - FN). Untuk *hyperparamter* pada target *Schiller* ini mendapat *hyperparameter* terbaik yaitu *hyperparameter* C 0.1 dengan akurasi 90%.

Berikut hasil perhitungan dari *accuracy, precision, recall*, dan *f1score* pada target *Cytology* yang didapatkan dari evaluasi *confusion matrix*.

Tabel 4. 54 *Confusion Matrix* Target *Cytology* Rasio 80:20

	Data Prediksi	
Data Aktual	**T**	**F**
T	0	5
F	0	167

Pada tabel 4.54 ditunjukkan hasil dari klasifikasi model untuk target *Citology*. Tidak ada sampel yang terdeteksi benar positif (*True Positives* - TP), 167 sampel yang terdeteksi benar negatif (*True Negatives* - TN), tidak ada sampel yang salah terdeteksi positif (*False Positives* - FP), dan 5 sampel positif yang salah terdeteksi negatif (*False Negatives* - FN). Untuk *hyperparamter* pada target

Citology ini mendapat *hyperparameter* terbaik yaitu *hyperparameter* C 0.1 dengan akurasi 96%.

Berikut hasil perhitungan dari *accuracy, precision, recall,* dan *f1score* pada target *Biopsy* yang didapatkan dari evaluasi *confusion matrix*.

Tabel 4. 55 *Confusion Matrix* Target *Biopsy* Rasio 80:20

	Data Prediksi	
Data Aktual	**T**	**F**
T	1	10
F	0	161

Pada tabel 4.47 ditunjukkan hasil dari klasifikasi model untuk target *Biopsy*. 1 sampel yang terdeteksi benar positif (*True Positives* - TP), 161 sampel yang terdeteksi benar negatif (*True Negatives* - TN), tidak ada sampel yang salah terdeteksi positif (*False Positives* - FP), dan 10 sampel positif yang salah terdeteksi negatif (*False Negatives* - FN). Untuk *hyperparamter* pada target *Biopsy* ini mendapat *hyperparameter* terbaik yaitu *hyperparameter* C 0.1 dengan akurasi 94%.

4.3.3.3 Pengujian Model C

Pada pengujian model C, nilai performa dari sistem meliputi *accuracy, precision recall,* dan *f1score* yang ditunjukkan pada tabel 4.56 dengan rasio perbandingan data latih dan data uji sebesar 70:30. Pada pengujian ini akan digunakan dalam empat target yaitu *Hinselmann, Schiller, Cytology*, dan *Biopsy*. Dengan pengujian 70:30 ini hasil akurasi paling tinggi didapatkan oleh target *Hinselmann* dengan nilai 76%.

Tabel 4. 56 Performa Sistem Rasio 70:30

Target	**Akurasi**	**Presisi**	**Recall**	**f1-Score**
Hinselmann	96%	0%	0%	0%
Schiller	92%	100%	4%	9%
Cytology	96%	0%	0%	0%
Biopsy	94%	100%	6%	12%

Hasil perhitungan *accuracy, precision, recall*, dan *f1score* pada target *Hinselmann* yang didapatkan dari evaluasi *confusion matrix* sebagai berikut.

Tabel 4. 57 *Confusion Matrix* Target *Hinselmann* Rasio 70:30

	Data Prediksi	
Data Aktual	**T**	**F**
T	0	10
F	0	248

Pada tabel 4.57 ditunjukkan hasil dari klasifikasi model untuk target *Hinselmann*. Tidak ada sampel yang terdeteksi benar positif (*True Positives* - TP), 248 sampel yang terdeteksi benar negatif (*True Negatives* - TN), tidak ada sampel yang salah terdeteksi positif (*False Positives* - FP), dan 10 sampel positif yang salah terdeteksi negatif (*False Negatives* - FN). Untuk *hyperparamter* pada target *Hinselmann* ini mendapat *hyperparameter* terbaik yaitu *hyperparameter* C 0.1 dengan akurasi 96%.

Berikut hasil perhitungan dari *accuracy, precision, recall*, dan *f1score* pada target *Schiller* yang didapatkan dari evaluasi *confusion matrix*.

Tabel 4. 58 *Confusion Matrix* Target *Schiller* Rasio 70:30

	Data Prediksi	
Data Aktual	**T**	**F**
T	1	20
F	0	237

Pada tabel 4.58 ditunjukkan hasil dari klasifikasi model untuk target *Schiller*. 1 sampel yang terdeteksi benar positif (*True Positives* - TP), 237 sampel yang terdeteksi benar negatif (*True Negatives* - TN), tidak ada sampel yang salah terdeteksi positif (*False Positives* - FP), dan 20 sampel positif yang salah terdeteksi negatif (*False Negatives* - FN).Untuk *hyperparamter* pada target *Schilerr* ini mendapat *hyperparameter* terbaik yaitu *hyperparameter* C 0.1 dengan akurasi 92%.

Berikut hasil perhitungan dari *accuracy, precision, recall,* dan *f1score* pada target *Cytology* yang didapatkan dari evaluasi *confusion matrix* .

Tabel 4. 59 *Confusion Matrix* Target *Cytology* Rasio 70:30

	Data Prediksi	
Data Aktual	**T**	**F**
T	0	8
F	0	250

Pada tabel 4.59 ditunjukkan hasil dari klasifikasi model untuk target *Citology*. Tidak ada sampel yang terdeteksi benar positif (*True Positives* - TP), 250 sampel yang terdeteksi benar negatif (*True Negatives* - TN), tidak ada sampel yang salah terdeteksi positif (*False Positives* - FP), dan 8 sampel positif yang salah terdeteksi negatif (*False Negatives* - FN). Untuk *hyperparamter* pada target *Citology* ini mendapat *hyperparameter* terbaik yaitu *hyperparameter* C 0.1 dengan akurasi 96%.

Berikut hasil perhitungan dari *accuracy, precision, recall,* dan *f1score* pada target *Biopsy* yang didapatkan dari evaluasi *confusion matrix*.

Tabel 4. 60 *Confusion Matrix* Target *Biopsy* Rasio 70:30

	Data Prediksi	
Data Aktual	**T**	**F**
T	1	14
F	0	243

Pada tabel 4.60 ditunjukkan hasil dari klasifikasi model untuk target *Biopsy*. 1 sampel yang terdeteksi benar positif (*True Positives* - TP), 243 sampel yang terdeteksi benar negatif (*True Negatives* - TN), tidak ada sampel yang salah terdeteksi positif (*False Positives* - FP), dan 14 sampel positif yang salah terdeteksi negatif (*False Negatives* - FN). Untuk *hyperparamter* pada target *Biopsy* ini mendapat *hyperparameter* terbaik yaitu *hyperparameter* C 0.1 dengan akurasi 94%.

4.3.3.4 Pengujian Model D

Pada pengujian model D, nilai performa dari sistem meliputi *accuracy*, *precision recall*, dan *f1score* yang ditunjukkan pada tabel 4.61 dengan rasio perbandingan data latih dan data uji sebesar 60:40. Pada pengujian ini akan digunakan dalam empat target yaitu *Hinselmann*, *Schiller*, *Cytology*, dan *Biopsy*. Dengan pengujian 60:40 ini hasil akurasi paling tinggi didapatkan oleh target *Hinselmann* dengan nilai 96%.

Tabel 4. 61 Performa Sistem Rasio 60:40

Target	Akurasi	Presisi	Recall	f1-Score
Hinselmann	96%	0%	0%	0%
Schiller	92%	100%	3%	6%
Cytology	95%	0%	0%	0%
Biopsy	93%	100%	4%	8%

Hasil perhitungan *accuracy, precision, recall*, dan *f1score* pada target *Hinselmann* yang didapatkan dari evaluasi *confusion matrix* sebagai berikut.

Tabel 4. 62 *Confusion Matrix* Target *Hinselmann* Rasio 60:40

	Data Prediksi	
Data Aktual	**T**	**F**
T	0	13
F	0	331

Pada tabel 4.62 ditunjukkan hasil dari klasifikasi model untuk target *Hinselmann*. Tidak ada sampel yang terdeteksi benar positif (*True Positives* - TP), 331 sampel yang terdeteksi benar negatif (*True Negatives* - TN), tidak ada sampel yang salah terdeteksi positif (*False Positives* - FP), dan 13 sampel positif yang salah terdeteksi negatif (*False Negatives* - FN). Untuk *hyperparamter* pada target *Hinselmann* ini mendapat *hyperparameter* terbaik yaitu *hyperparameter* C 0.1 dengan akurasi 96%.

Berikut hasil perhitungan dari *accuracy, precision, recall,* dan *f1score* pada target *Schiller* yang didapatkan dari evaluasi *confusion matrix*.

Tabel 4. 63 Confusion Matrix Target Schiller Rasio 60:40

	Data Prediksi	
Data Aktual	**T**	**F**
T	1	27
F	0	316

Pada tabel 4.63 ditunjukkan hasil dari klasifikasi model untuk target *Schiller*. 1 sampel yang terdeteksi benar positif (*True Positives* - TP), 316 sampel yang terdeteksi benar negatif (*True Negatives* - TN), tidak ada sampel yang salah terdeteksi positif (*False Positives* - FP), dan 27 sampel positif yang salah terdeteksi negatif (*False Negatives* - FN). Untuk *hyperparamter* pada target *Schiller* ini mendapat *hyperparameter* terbaik yaitu *hyperparameter* C 0.1 dengan akurasi 92%.

Berikut hasil perhitungan dari *accuracy, precision, recall,* dan *f1score* pada target *Cytology* yang didapatkan dari evaluasi *confusion matrix*.

Tabel 4. 64 *Confusion Matrix* Target *Cytology* Rasio 60:40

	Data Prediksi	
Data Aktual	**T**	**F**
T	0	16
F	0	328

Pada tabel 4.64 ditunjukkan hasil dari klasifikasi model untuk target *Citology*. Tidak ada sampel yang terdeteksi benar positif (*True Positives* - TP), 328 sampel yang terdeteksi benar negatif (*True Negatives* - TN), tidak ada sampel yang salah terdeteksi positif (*False Positives* - FP), dan 16 sampel positif yang salah terdeteksi negatif (*False Negatives* - FN). Untuk *hyperparamter* pada target *Citology* ini mendapat *hyperparameter* terbaik yaitu *hyperparameter* C 0.1 dengan akurasi 95%.

Berikut hasil perhitungan dari *accuracy, precision, recall,* dan *f1score* pada target *Biopsy* yang didapatkan dari evaluasi *confusion matrix*.

Tabel 4. 65 *Confusion Matrix* Target *Biopsy* Rasio 60:40

	Data Prediksi	
Data Aktual	**T**	**F**
T	1	21
F	0	322

Pada tabel 4.65 ditunjukkan hasil dari klasifikasi model untuk target *Biopsy*. 1 sampel yang terdeteksi benar positif (*True Positives* - TP), 322 sampel yang terdeteksi benar negatif (*True Negatives* - TN), tidak ada sampel yang salah terdeteksi positif (*False Positives* - FP), dan 322 sampel positif yang salah terdeteksi negatif (*False Negatives* - FN). Untuk *hyperparamter* pada target *Biopsy* ini mendapat *hyperparameter* terbaik yaitu *hyperparameter* C 0.1 dengan akurasi 93%.

4.3.4 Pengujian dengan Metode SVM tanpa PCA SMOTE

Pada pengujian ini, akan dilakukan tanpa menggunakan teknik PCA (*Principal Component Analysis*) dan juga tidak menggunakan teknik SMOTE (*Synthetic Minority Over-sampling Technique*) untuk menangani ketidakseimbangan kelas dalam dataset.

4.3.4.1 Pengujian Model A

Pada pengujian model A, nilai performa dari sistem meliputi *accuracy, precision recall,* dan *f1score* yang ditunjukkan pada tabel 4.66 dengan rasio perbandingan data latih dan data uji sebesar 90:10. Pada pengujian ini akan digunakan dalam empat target yaitu *Hinselmann, Schiller, Cytology,* dan *Biopsy*.

Dengan pengujian 90:10 ini hasil akurasi paling tinggi didapatkan oleh target *Citology* dengan nilai 98%.

Tabel 4. 66 Tabel 4. 46 Performa Sistem Rasio 90:10

Target	**Akurasi**	**Presisi**	**Recall**	**f1-Score**
Hinselmann	97%	0%	0%	0%
Schiller	94%	0%	0%	0%
Cytology	98%	0%	0%	0%
Biopsy	97%	0%	0%	0%

Hasil perhitungan *accuracy, precision, recall,* dan *f1score* pada target *Hinselmann* yang didapatkan dari evaluasi *confusion matrix* sebagai berikut.

Tabel 4. 67 *Confusion Matrix* Target *Hinselmann* Rasio 90:10 dengan Metode SVM PCA tanpa SMOTE

	Data Prediksi	
Data Aktual	**T**	**F**
T	0	2
F	0	84

Pada tabel 4.67 ditunjukkan hasil dari klasifikasi model untuk target *Hinselmann*. Tidak ada sampel yang terdeteksi benar positif (*True Positives* - TP), 84 sampel yang terdeteksi benar negatif (*True Negatives* - TN), tidak ada sampel yang salah terdeteksi positif (*False Positives* - FP), dan 2 sampel positif yang salah terdeteksi negatif (*False Negatives* - FN). Untuk *hyperparamter* pada target *Hinselmann* ini mendapat *hyperparameter* terbaik yaitu *hyperparameter* C 0.1 dengan akurasi 96%.

Berikut hasil perhitungan dari *accuracy, precision, recall,* dan *f1score* pada target *Schiller* yang didapatkan dari evaluasi *confusion matrix*.

Tabel 4. 68 *Confusion Matrix* Target *Schiller* Rasio 90:10

	Data Prediksi	
Data Aktual	**T**	**F**
T	0	5
F	0	81

Pada tabel 4.68 ditunjukkan hasil dari klasifikasi model untuk target *Schiller*. Tidak ada sampel yang terdeteksi benar positif (*True Positives* - TP), 81 sampel yang terdeteksi benar negatif (*True Negatives* - TN), tidak ada sampel yang salah terdeteksi positif (*False Positives* - FP), dan 5 sampel positif yang salah terdeteksi negatif (*False Negatives* - FN). Untuk *hyperparamter* pada target *Schiller* ini mendapat *hyperparameter* terbaik yaitu *hyperparameter* C 0.1 dengan akurasi 94%.

Berikut hasil perhitungan dari *accuracy, precision, recall*, dan *f1score* pada target *Cytology* yang didapatkan dari evaluasi *confusion matrix*.

Tabel 4. 69 *Confusion Matrix* Target *Cytology* Rasio 90:10

	Data Prediksi	
Data Aktual	**T**	**F**
T	0	1
F	0	85

Pada tabel 4.69 ditunjukkan hasil dari klasifikasi model untuk target *Citology*. Tidak ada sampel yang terdeteksi benar positif (*True Positives* - TP), 85 sampel yang terdeteksi benar negatif (*True Negatives* - TN), tidak ada sampel yang salah terdeteksi positif (*False Positives* - FP), dan 1 sampel positif yang salah terdeteksi negatif (*False Negatives* - FN). Untuk *hyperparamter* pada target *Citology* ini mendapat *hyperparameter* terbaik yaitu *hyperparameter* C 0.1 dengan akurasi 98%.

Berikut hasil perhitungan dari *accuracy, precision, recall*, dan *f1score* pada target *Biopsy* yang didapatkan dari evaluasi *confusion matrix*.

Tabel 4. 70 *Confusion Matrix* Target *Biopsy* Rasio 90:10

	Data Prediksi	
Data Aktual	**T**	**F**
T	0	2
F	0	84

Pada tabel 4.70 ditunjukkan hasil dari klasifikasi model untuk target *Biopsy*. Tidak ada sampel yang terdeteksi benar positif (*True Positives* - TP), 84 sampel yang terdeteksi benar negatif (*True Negatives* - TN), tidak ada sampel yang salah terdeteksi positif (*False Positives* - FP), dan 2 sampel positif yang salah terdeteksi negatif (*False Negatives* - FN). Untuk *hyperparamter* pada target *Biopsy* ini mendapat *hyperparameter* terbaik yaitu *hyperparameter* C 0.1 dengan akurasi 97%.

4.3.4.2 Pengujian Model B

Pada pengujian model B, nilai performa dari sistem meliputi *accuracy, precision recall,* dan *f1score* yang ditunjukkan pada tabel 4.71 dengan rasio perbandingan data latih dan data uji sebesar 80:20. Pada pengujian ini akan digunakan dalam empat target yaitu *Hinselmann, Schiller, Cytology,* dan *Biopsy*. Dengan pengujian 80:20 ini hasil akurasi paling tinggi didapatkan oleh target *Citology* dengan nilai 78%.

Tabel 4. 71 Performa Sistem Rasio 80:20

Target	Akurasi	Presisi	Recall	f1-Score
Hinselmann	95%	0%	0%	0%
Schiller	90%	0%	0%	0%
Cytology	98%	0%	0%	0%
Biopsy	94%	100%	6%	16%

Berikut hasil perhitungan *accuracy, precision, recall,* dan *f1score* pada target *Hinselmann* yang didapatkan dari evaluasi *confusion matrix*.

Tabel 4. 72 *Confusion Matrix* Target *Hinselmann* Rasio 80:20

	Data Prediksi	
Data Aktual	**T**	**F**
T	0	8
F	0	164

Pada tabel 4.72 ditunjukkan hasil dari klasifikasi model untuk target *Hinselmann*. Tidak ada sampel yang terdeteksi benar positif (*True Positives* - TP), 164 sampel yang terdeteksi benar negatif (*True Negatives* - TN), tidak ada sampel yang salah terdeteksi positif (*False Positives* - FP), dan 8 sampel positif yang salah terdeteksi negatif (*False Negatives* - FN). Untuk *hyperparamter* pada target *Hinselmann* ini mendapat *hyperparameter* terbaik yaitu *hyperparameter* C 0.1 dengan akurasi 95%.

Berikut hasil perhitungan dari *accuracy, precision, recall*, dan *f1score* pada target *Schiller* yang didapatkan dari evaluasi *confusion matrix*.

Tabel 4. 73 *Confusion Matrix* Target *Schiller* Rasio 80:20

	Data Prediksi	
Data Aktual	**T**	**F**
T	0	17
F	0	155

Pada tabel 4.73 ditunjukkan hasil dari klasifikasi model untuk target *Schiller*. Tidak ada sampel yang terdeteksi benar positif (*True Positives* - TP), 155 sampel yang terdeteksi benar negatif (*True Negatives* - TN), tidak ada sampel yang salah terdeteksi positif (*False Positives* - FP), dan 17 sampel positif yang salah terdeteksi negatif (*False Negatives* - FN). Untuk *hyperparamter* pada target *Schiller* ini mendapat *hyperparameter* terbaik yaitu *hyperparameter* C 0.1 dengan akurasi 90%.

Berikut hasil perhitungan dari *accuracy, precision, recall*, dan *f1score* pada target *Cytology* yang didapatkan dari evaluasi *confusion matrix* .

Tabel 4. 74 *Confusion Matrix* Target *Cytology* Rasio 80:20

	Data Prediksi	
Data Aktual	**T**	**F**
T	0	5
F	0	167

Pada tabel 4.74 ditunjukkan hasil dari klasifikasi model untuk target *Citology*. Tidak ada sampel yang terdeteksi benar positif (*True Positives* - TP), 167 sampel yang terdeteksi benar negatif (*True Negatives* - TN), tidak ada sampel yang salah terdeteksi positif (*False Positives* - FP), dan 5 sampel positif yang salah terdeteksi negatif (*False Negatives* - FN). Untuk *hyperparamter* pada target *Citology* ini mendapat *hyperparameter* terbaik yaitu *hyperparameter* C 0.1 dengan akurasi 96%.

Berikut hasil perhitungan dari *accuracy, precision, recall*, dan *f1score* pada target *Biopsy* yang didapatkan dari evaluasi *confusion matrix*.

Tabel 4. 75 *Confusion Matrix* Target *Biopsy* Rasio 80:20

	Data Prediksi	
Data Aktual	**T**	**F**
T	1	10
F	0	161

Pada tabel 4.75 ditunjukkan hasil dari klasifikasi model untuk target *Biopsy*. 1 sampel yang terdeteksi benar positif (*True Positives* - TP), 161 sampel yang terdeteksi benar negatif (*True Negatives* - TN), tidak ada sampel yang salah terdeteksi positif (*False Positives* - FP), dan 10 sampel positif yang salah terdeteksi negatif (*False Negatives* - FN). Untuk *hyperparamter* pada target *Biopsy* ini mendapat *hyperparameter* terbaik yaitu *hyperparameter* C 0.1 dengan akurasi 94%.

4.3.4.3 Pengujian Model C

Pada pengujian model C, nilai performa dari sistem meliputi *accuracy, precision recall*, dan *f1score* yang ditunjukkan pada tabel 4.76 dengan rasio perbandingan data latih dan data uji sebesar 70:30. Pada pengujian ini akan digunakan dalam empat target yaitu *Hinselmann, Schiller, Cytology*, dan *Biopsy*.

Dengan pengujian 70:30 ini hasil akurasi paling tinggi didapatkan oleh target *Hinselmann* dengan nilai 76%.

Tabel 4. 76 Performa Sistem Rasio 70:30

Target	Akurasi	Presisi	Recall	f1-Score
Hinselmann	96%	0%	0%	0%
Schiller	92%	100%	4%	9%
Cytology	96%	0%	0%	0%
Biopsy	94%	100%	6%	12%

Hasil perhitungan *accuracy, precision, recall*, dan *f1score* pada target *Hinselmann* yang didapatkan dari evaluasi *confusion matrix* sebagai berikut.

Tabel 4. 77 *Confusion Matrix* Target *Hinselmann* Rasio 70:30

	Data Prediksi	
Data Aktual	**T**	**F**
T	0	10
F	0	248

Pada tabel 4.77 ditunjukkan hasil dari klasifikasi model untuk target *Hinselmann*. Tidak ada sampel yang terdeteksi benar positif (*True Positives* - TP), 248 sampel yang terdeteksi benar negatif (*True Negatives* - TN), tidak ada sampel yang salah terdeteksi positif (*False Positives* - FP), dan 10 sampel positif yang salah terdeteksi negatif (*False Negatives* - FN). Untuk *hyperparamter* pada target *Hinselmann* ini mendapat *hyperparameter* terbaik yaitu *hyperparameter* C 0.1 dengan akurasi 96%.

Berikut hasil perhitungan dari *accuracy, precision, recall*, dan *f1score* pada target *Schiller* yang didapatkan dari evaluasi *confusion matrix*.

Tabel 4. 78 *Confusion Matrix* Target *Schiller* Rasio 70:30

	Data Prediksi	
Data Aktual	**T**	**F**
T	1	20
F	0	237

Pada tabel 4.78 ditunjukkan hasil dari klasifikasi model untuk target *Schiller*. 1 sampel yang terdeteksi benar positif (*True Positives* - TP), 237 sampel yang terdeteksi benar negatif (*True Negatives* - TN), tidak ada sampel yang salah terdeteksi positif (*False Positives* - FP), dan 20 sampel positif yang salah terdeteksi negatif (*False Negatives* - FN).Untuk *hyperparamter* pada target *Schilerr* ini mendapat *hyperparameter* terbaik yaitu *hyperparameter* C 0.1 dengan akurasi 92%.

Berikut hasil perhitungan dari *accuracy, precision, recall*, dan *f1score* pada target *Cytology* yang didapatkan dari evaluasi *confusion matrix*.

Tabel 4. 79 *Confusion Matrix* Target *Cytology* Rasio 70:30

	Data Prediksi	
Data Aktual	**T**	**F**
T	0	8
F	0	250

Pada tabel 4.79 ditunjukkan hasil dari klasifikasi model untuk target *Citology*. Tidak ada sampel yang terdeteksi benar positif (*True Positives* - TP), 250 sampel yang terdeteksi benar negatif (*True Negatives* - TN), tidak ada sampel yang salah terdeteksi positif (*False Positives* - FP), dan 8 sampel positif yang salah terdeteksi negatif (*False Negatives* - FN). Untuk *hyperparamter* pada target *Citology* ini mendapat *hyperparameter* terbaik yaitu *hyperparameter* C 0.1 dengan akurasi 96%.

Berikut hasil perhitungan dari *accuracy, precision, recall*, dan *f1score* pada target *Biopsy* yang didapatkan dari evaluasi *confusion matrix*.

Tabel 4. 80 *Confusion Matrix* Target *Biopsy* Rasio 70:30

	Data Prediksi	
Data Aktual	**T**	**F**
T	1	14
F	0	243

Pada tabel 4.80 ditunjukkan hasil dari klasifikasi model untuk target *Biopsy*. 1 sampel yang terdeteksi benar positif (*True Positives* - TP), 243 sampel yang terdeteksi benar negatif (*True Negatives* - TN), tidak ada sampel yang salah terdeteksi positif (*False Positives* - FP), dan 14 sampel positif yang salah terdeteksi negatif (*False Negatives* - FN). Untuk *hyperparamter* pada target *Biopsy* ini mendapat *hyperparameter* terbaik yaitu *hyperparameter* C 0.1 dengan akurasi 94%.

4.3.4.4 Pengujian Model D

Pada pengujian model D, nilai performa dari sistem meliputi *accuracy, precision recall,* dan *f1score* yang ditunjukkan pada tabel 4.81 dengan rasio perbandingan data latih dan data uji sebesar 60:40. Pada pengujian ini akan digunakan dalam empat target yaitu *Hinselmann, Schiller, Cytology,* dan *Biopsy*. Dengan pengujian 60:40 ini hasil akurasi paling tinggi didapatkan oleh target *Hinselmann* dengan nilai 96%.

Tabel 4. 81 Performa Sistem Rasio 60:40

Target	Akurasi	Presisi	Recall	f1-Score
Hinselmann	96%	0%	0%	0%
Schiller	92%	100%	3%	6%
Cytology	95%	0%	0%	0%
Biopsy	93%	100%	4%	8%

Hasil perhitungan *accuracy, precision, recall*, dan *f1score* pada target *Hinselmann* yang didapatkan dari evaluasi *confusion matrix* sebagai berikut.

Tabel 4. 82 *Confusion Matrix* Target *Hinselmann* Rasio 60:40

	Data Prediksi	
Data Aktual	**T**	**F**
T	0	13
F	0	331

Pada tabel 4.82 ditunjukkan hasil dari klasifikasi model untuk target *Hinselmann*. Tidak ada sampel yang terdeteksi benar positif (*True Positives* - TP), 331 sampel yang terdeteksi benar negatif (*True Negatives* - TN), tidak ada sampel yang salah terdeteksi positif (*False Positives* - FP), dan 13 sampel positif yang salah terdeteksi negatif (*False Negatives* - FN). Untuk *hyperparamter* pada target *Hinselmann* ini mendapat *hyperparameter* terbaik yaitu *hyperparameter* C 0.1 dengan akurasi 96%.

Berikut hasil perhitungan dari *accuracy, precision, recall*, dan *f1score* pada target *Schiller* yang didapatkan dari evaluasi *confusion matrix*.

Tabel 4. 83 Confusion Matrix Target Schiller Rasio 60:40

	Data Prediksi	
Data Aktual	**T**	**F**
T	1	27
F	0	316

Pada tabel 4.83 ditunjukkan hasil dari klasifikasi model untuk target *Schiller*. 1 sampel yang terdeteksi benar positif (*True Positives* - TP), 316 sampel yang terdeteksi benar negatif (*True Negatives* - TN), tidak ada sampel yang salah terdeteksi positif (*False Positives* - FP), dan 27 sampel positif yang salah terdeteksi negatif (*False Negatives* - FN). Untuk *hyperparamter* pada target *Schiller* ini mendapat *hyperparameter* terbaik yaitu *hyperparameter* C 0.1 dengan akurasi 92%.

Berikut hasil perhitungan dari *accuracy, precision, recall*, dan *f1score* pada target *Cytology* yang didapatkan dari evaluasi *confusion matrix*.

Tabel 4. 84 *Confusion Matrix* Target *Cytology* Rasio 60:40

	Data Prediksi	
Data Aktual	**T**	**F**
T	0	16
F	0	328

Pada tabel 4.84 ditunjukkan hasil dari klasifikasi model untuk target *Citology*. Tidak ada sampel yang terdeteksi benar positif (*True Positives* - TP), 328 sampel yang terdeteksi benar negatif (*True Negatives* - TN), tidak ada sampel yang salah terdeteksi positif (*False Positives* - FP), dan 16 sampel positif yang salah terdeteksi negatif (*False Negatives* - FN). Untuk *hyperparamter* pada target *Citology* ini mendapat *hyperparameter* terbaik yaitu *hyperparameter* C 0.1 dengan akurasi 95%.

Berikut hasil perhitungan dari *accuracy, precision, recall*, dan *f1score* pada target *Biopsy* yang didapatkan dari evaluasi *confusion matrix*.

Tabel 4. 85 *Confusion Matrix* Target *Biopsy* Rasio 60:40

	Data Prediksi	
Data Aktual	**T**	**F**
T	1	21
F	0	322

Pada tabel 4.85 ditunjukkan hasil dari klasifikasi model untuk target *Biopsy*. 1 sampel yang terdeteksi benar positif (*True Positives* - TP), 322 sampel yang terdeteksi benar negatif (*True Negatives* - TN), tidak ada sampel yang salah terdeteksi positif (*False Positives* - FP), dan 322 sampel positif yang salah terdeteksi negatif (*False Negatives* - FN). Untuk *hyperparamter* pada target *Biopsy* ini mendapat *hyperparameter* terbaik yaitu *hyperparameter* C 0.1 dengan akurasi 93%.

4.4 Pembahasan

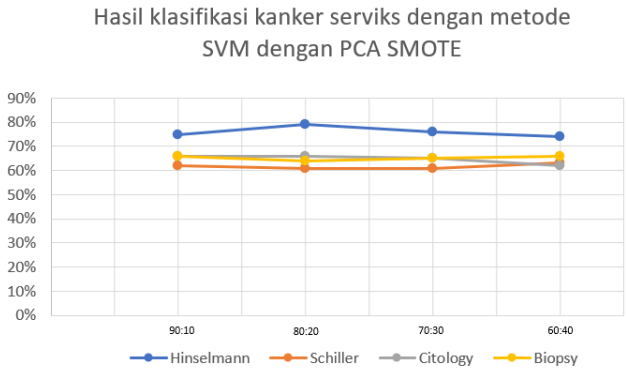
Pada penelitian ini, peneliti menerapkan teknik SMOTE (*Synthetic Minority Over-sampling Technique*) yang digunakan sebelum melakukan reduksi dimensi dengan metode PCA (*Principal Component Analysis)*. Terdapat empat target dalam

penelitian ini yaitu *Hinselmann, Schiller*, *Cytology*, dan *Biopsy*. Setiap target memiliki distribusi data positif dan negatif yang berbeda, jadi penerapan SMOTE bertujuan untuk menyeimbangkan jumlah sampel pada masing-masing kelas. Setelah proses penyeimbangan data, PCA digunakan untuk mengurangi dimensi variabel dalam kumpulan data. Distribusi awal data positif dan negatif tidak sama untuk setiap target, jadi hasil komponen utama PCA tidak sama. Namun pada tiap target dari hasil PCA mendapatkan sama yaitu PC1-PC19 yang mendapatkan nilai variance rasio yang melebihi 1 dan cumulative varaince rasio diatas 95% namun untuk nilai dari setiap PCnya berbeda.

Berdasarkan hasil dari pengujian yang sudah dijalankan dengan berbagai rasio perbandingan data latih dan data uji pada 4 target untuk menentukan model skenario yang paling optimal dalam melatih *Support Vector Machine* (SVM) berbasis *Principal Component Analysis* (PCA) dalam mengklasifikasikan kanker serviks. Berikut merupakan hasil *accuracy* dari setiap targetnya.

Tabel 4. 86 Hasil Akurasi kanker serviks dengan metode SVM dengan PCA SMOTE

Model	Rasio Data	Target	Accuracy
A	90:10	*Hinselmann*	75%
		Schiller	62%
		Cytology	66%
		Biopsy	66%
B	80:20	*Hinselmann*	79%
		Schiller	61%
		Cytology	66%
		Biopsy	64%
C	70:30	*Hinselmann*	76%
		Schiller	61%
		Cytology	65%
		Biopsy	65%
D	60:30	*Hinselmann*	74%
		Schiller	63%
		Cytology	62%
		Biopsy	66%



Gambar 4. 9 Grafik Hasil Klasifikasi dengan Metode SVM dengan PCA SMOTE

Pada gambar 4.9 menunjukkan bahwa hasil akurasi pada klasifikasi kanker serviks menggunakan metode SVM berbasis PCA mendapatkan hasil akurasi paling tinggi pada target *Hinselmann* dengan berbagai macam rasio perbandingan. Nilai paling tinggi didapatkan 79% dari rasio perbadingan data latih dan data uji 80:20.

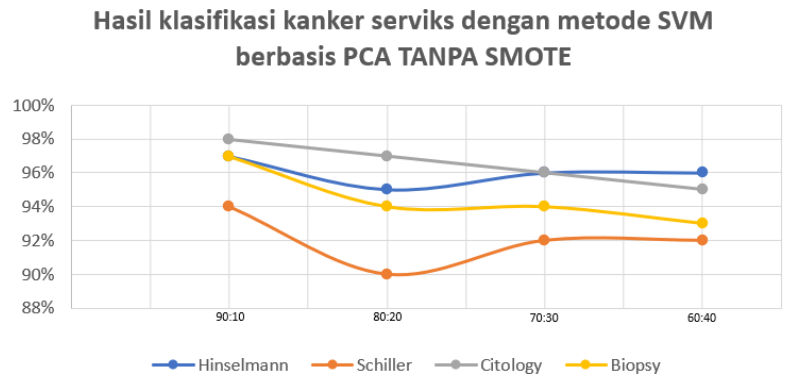
Pada tabel 4.86, terlihat bahwa penerapan SMOTE sebelum melakukan reduksi dimensi dengan PCA berdampak pada hasil akhir klasifikasi untuk target *Hinselmann*, *Schiller*, *Cytology*, dan *Biopsy*. Meskipun terdapat variasi dalam distribusi awal data positif dan negatif, hasil akurasi menunjukkan bahwa perubahan proporsi data latih dan uji dari 90:10, 80:20, 70:30, hingga 60:40 tidak memberikan perbedaan yang signifikan dalam kinerja model.

Berikut merupakan hasil *accuracy* dari pengujian dengan SVM PCA tanpa SMOTE.

Tabel 4. 87 Hasil Pengujian tanpa PCA

Model	Rasio Data	Target	Accuracy
A	90:10	*Hinselmann*	97%
		Schiller	94%
		Cytology	98%
		Biopsy	97%
B	80:10	*Hinselmann*	95%
		Schiller	90%
		Cytology	97%
		Biopsy	94%

C	70:30	*Hinselmann*	95%
		Schiller	90%
		Cytology	97%
		Biopsy	94%
D	60:30	*Hinselmann*	95%
		Schiller	90%
		Cytology	97%
		Biopsy	94%



Gambar 4. 10 Grafik Hasil Klasifikasi dengan Metode SVM tanpa PCA

Pada gambar 4.10 menunjukkan bahwa hasil akurasi pada klasifikasi kanker serviks menggunakan metode SVM tanpa PCA mendapatkan hasil akurasi paling tinggi pada target *Hinselmann* dengan berbagai macam rasio perbandingan. Nilai paling tinggi didapatkan 98% dari rasio perbadingan data latih dan data uji 90:10.

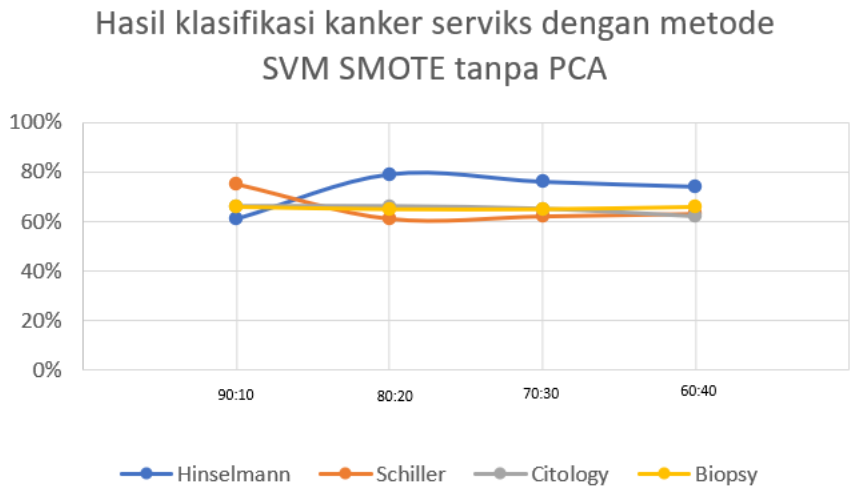
Pada tabel 4.87, hasil klasifikasi yang dihasilkan oleh tanpa penerapan SMOTE dan menerapkan dengan PCA menunjukkan akurasi yang sangat tinggi untuk target Hinselmann. Namun, hasil ini juga menunjukkan kemungkinan *overfitting*, karena akurasi mendekati 100% biasanya menunjukkan bahwa model mungkin terlalu menyesuaikan dengan data yang dilatih. Hasil ini menunjukkan bahwa SMOTE tidak digunakan untuk menyeimbangkan distribusi kelas meskipun PCA digunakan untuk mengurangi dimensi data. Sempurnanya akurasi ini

menunjukkan bahwa model mengalami *overfitting* yang berarti bahwa model sangat menyesuaikan dengan data latih dan bisa dikatakan tidak akan bekerja dengan baik pada data uji yang belum pernah dilihat sebelumnya.

Berikut merupakan hasil *accuracy* dari pengujian dengan SVM SMOTE tanpa PCA.

Tabel 4. 88 Hasil Akurasi kanker serviks dengan metode SVM SMOTE tanpa PCA

Model	Rasio Data	Target	Accuracy
A	90:10	*Hinselmann*	61%
		Schiller	75%
		Cytology	66%
		Biopsy	66%
B	80:20	*Hinselmann*	79%
		Schiller	61%
		Cytology	66%
		Biopsy	65%
C	70:30	*Hinselmann*	76%
		Schiller	62%
		Cytology	65%
		Biopsy	65%
D	60:30	*Hinselmann*	74%
		Schiller	63%
		Cytology	62%
		Biopsy	66%



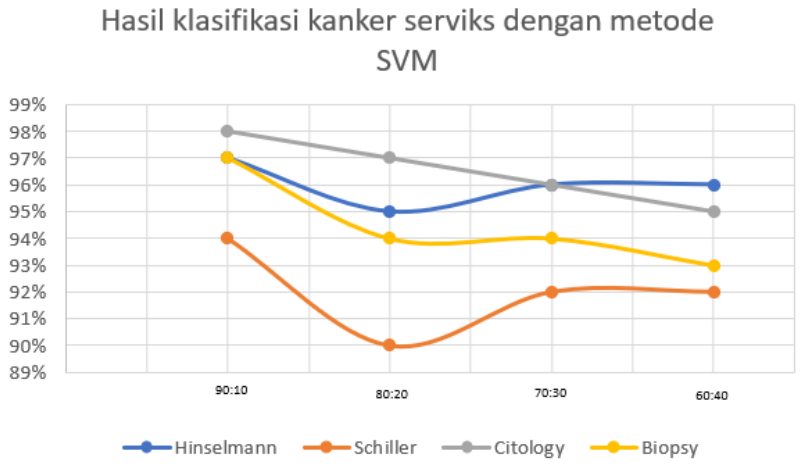
Gambar 4. 11 Grafik Hasil Klasifikasi dengan Metode SVM SMOTE tanpa PCA

Pada gambar 4.11 menunjukkan bahwa hasil akurasi pada klasifikasi kanker serviks menggunakan metode SVM berbasis PCA mendapatkan hasil akurasi paling tinggi pada target *Hinselmann* dengan berbagai macam rasio perbandingan. Nilai paling tinggi didapatkan 79% dari rasio perbadingan data latih dan data uji 80:20.

Berikut merupakan hasil *accuracy* dari pengujian dengan SVM tanpa PCA SMOTE.

Tabel 4. 89 Hasil Pengujian tanpa PCA SMOTE

Model	Rasio Data	Target	Accuracy
A	90:10	*Hinselmann*	97%
		Schiller	94%
		Cytology	98%
		Biopsy	97%
B	80:10	*Hinselmann*	95%
		Schiller	90%
		Cytology	97%
		Biopsy	94%
C	70:30	*Hinselmann*	95%
		Schiller	90%
		Cytology	97%
		Biopsy	94%
D	60:30	*Hinselmann*	95%
		Schiller	90%
		Cytology	97%
		Biopsy	94%



Gambar 4. 12 Grafik Hasil Klasifikasi dengan Metode SVM tanpa PCA SMOTE

Pada gambar 4.12 menunjukkan bahwa hasil akurasi pada klasifikasi kanker serviks menggunakan metode SVM tanpa PCA mendapatkan hasil akurasi paling tinggi pada target *Hinselmann* dengan berbagai macam rasio perbandingan. Nilai paling tinggi didapatkan 98% dari rasio perbadingan data latih dan data uji 90:10.

Pada tabel 4.89, hasil klasifikasi yang dihasilkan oleh tanpa penerapan SMOTE dan tanpa penerapan PCA menunjukkan akurasi yang sangat tinggi untuk target *Hinselmann*. Namun, hasil ini juga menunjukkan kemungkinan *overfitting*, karena akurasi mendekati 100% biasanya menunjukkan bahwa model mungkin terlalu menyesuaikan dengan data yang dilatih. Hasil ini menunjukkan bahwa SMOTE tidak digunakan untuk menyeimbangkan distribusi kelas meskipun PCA digunakan untuk mengurangi dimensi data. Sempurnanya akurasi ini menunjukkan bahwa model mengalami *overfitting* yang berarti bahwa model sangat menyesuaikan dengan data latih dan bisa dikatakan tidak akan bekerja dengan baik pada data uji yang belum pernah dilihat sebelumnya.

Berdasarkan hasil pengujian klasifikasi kanker serviks, baik dengan maupun tanpa menggunakan SMOTE dan PCA. Pertama, empat target (*Hinselmann*, *Schiller*, *Cytology*, dan *Biopsy*) menunjukkan akurasi yang cukup bervariasi setelah penggunaan SMOTE dan PCA. Target *Hinselmann* memiliki akurasi tertinggi dengan 79% untuk rasio data latih dan uji 80:20. Namun, akurasi ini tidak jauh berbeda dari hasil rasio perbandingan lainnya dan hal ini menunjukkan bahwa SMOTE dan PCA hanya menghasilkan peningkatan kecil dalam beberapa situasi.

PCA tanpa SMOTE menghasilkan akurasi yang sangat tinggi pada semua target, dengan nilai hampir 100%. Akurasi untuk target *Cytology* mencapai 98%

pada rasio data latih dan uji 90:10. Namun, akurasi yang terlalu tinggi ini menunjukkan kemungkinan *overfitting* dan hal ini terjadi ketika model terlalu menyesuaikan dengan data latih dan tidak berfungsi dengan baik pada data uji yang belum pernah dilihat sebelumnya. Ini menunjukkan bahwa meskipun PCA membantu mengurangi dimensi tanpa menyeimbangi data, model menjadi terlalu khusus untuk data latih.

Sebaliknya, hasil akurasi tetap menunjukkan perbedaan yang sama jika PCA tidak diterapkan tetapi SMOTE digunakan. Target *Hinselmann* juga memiliki akurasi tertinggi dengan 79% pada rasio data latih dan uji 80:20. Hal ini menunjukkan bahwa SMOTE memiliki efek yang lebih stabil tanpa PCA, yang berarti penyeimbangan data kelas mayoritas dan minoritas sebelum klasifikasi adalah langkah penting untuk meningkatkan keandalan dan stabilitas model.

Selain itu, hasil klasifikasi metode SVM tanpa PCA dan tanpa SMOTE menunjukkan akurasi yang sangat tinggi, hampir 100% untuk setiap target. Hasil ini sebanding dengan hasil SVM dengan PCA tanpa SMOTE, yang menunjukkan kemungkinan *overfitting* yang tinggi. *Overfitting* terjadi karena model sangat menyesuaikan dengan data latih, sehingga kinerja mungkin tidak dapat digeneralisasikan untuk data uji baru.

4.5 Integrasi Islam

Dalam Islam, pengetahuan dan aturan yang terstruktur sangat dihargai sebab dapat memudahkan dalam menjalani kehidupan yang sesuai dengan syariat-Nya. Al-Qur’an merupakan pedoman utama untuk hamba-Nya, yang mana memberikan

contoh tentang pentingnya klasifikasi dan pemahaman terorganisir. Seperti salah satu surat dalam Al-Qur'an yaitu Surat *An-Nisa'* ayat 23:

حُرِّمَتْ عَلَيْكُمْ أُمَّهَٰتُكُمْ وَبَنَاتُكُمْ وَأَخَوَٰتُكُمْ وَعَمَّٰتُكُمْ وَخَٰلَٰتُكُمْ وَبَنَاتُ ٱلْأَخِ وَبَنَاتُ ٱلْأُخْتِ وَأُمَّهَٰتُكُمُ ٱلَّٰتِىٓ أَرْضَعْنَكُمْ وَأَخَوَٰتُكُم مِّنَ ٱلرَّضَٰعَةِ وَأُمَّهَٰتُ نِسَآئِكُمْ وَرَبَٰٓئِبُكُمُ ٱلَّٰتِى فِى حُجُورِكُم مِّن نِّسَآئِكُمُ ٱلَّٰتِى دَخَلْتُم بِهِنَّ فَإِن لَّمْ تَكُونُوا۟ دَخَلْتُم بِهِنَّ فَلَا جُنَاحَ عَلَيْكُمْ وَحَلَٰٓئِلُ أَبْنَآئِكُمُ ٱلَّذِينَ مِنْ أَصْلَٰبِكُمْ وَأَن تَجْمَعُوا۟ بَيْنَ ٱلْأُخْتَيْنِ إِلَّا مَا قَدْ سَلَفَ ۗ إِنَّ ٱللَّهَ كَانَ غَفُورًا رَّحِيمًا

"Diharamkan atas kamu (mengawini) ibu-ibumu; anak-anakmu yang perempuan; saudara-saudaramu yang perempuan, saudara-saudara bapakmu yang perempuan; saudara-saudara ibumu yang perempuan; anak-anak perempuan dari saudara-saudaramu yang laki-laki; anak-anak perempuan dari saudara-saudaramu yang perempuan; ibu-ibumu yang menyusui kamu; saudara perempuan sepersusuan; ibu-ibu isterimu (mertua); anak-anak isterimu yang dalam pemeliharaanmu dari isteri yang telah kamu campuri, tetapi jika kamu belum campur dengan isterimu itu (dan sudah kamu ceraikan), maka tidak berdosa kamu mengawininya; (dan diharamkan bagimu) isteri-isteri anak kandungmu (menantu); dan menghimpunkan (dalam perkawinan) dua perempuan yang bersaudara, kecuali yang telah terjadi pada masa lampau; sesungguhnya Allah Maha Pengampun lagi Maha Penyayang."(QS. An-Nisa': 23)

Menurut *Tafsir Ibnu Katsir,* ayat ini menjelaskan bahwa Allah SWT dengan jelas telah mengklasifikasikan perempuan-perempuan yang haram dinikahi yang mana terdiri dari orang yang memiliki hubungan mahram melalui nasab, melalui hubungan persusuan, melalui hubungan perkawinan. Seperti halnya yang dikatakan oleh Ibnu Abbas, " Telah diharamkan untuk kalian tujuh hubungan nasab dan tujuh hubungan melalui perkawinan". Klasifikasi ini merupakan bagian dari kebijaksanaan Allah SWT untuk memastikan umat Islam agar lebih mudah dan jelas untuk memahami dan mematuhi aturan-aturan yang sudah ditetapkan. Hal ini mencerminkan pentingnya klasifikasi dalam memudahkan pemahaman dan penerapan aturan. Analoginya, dalam dunia medis yang terutama penanganan kanker serviks juga memainkan peran penting. Kanker serviks diklasifikasikan

berdasarkan tahapan perkembangan dari kanker tersebut, dimana hal itu sangat penting untuk menentukan prognosis dan rencana kedepan untuk pengobatan yang tepat bagi pasien.

Dalam surat *At-Thalaq* ayat 4, Al-Qur'an juga mengatakan tentang masa *iddah* wanita. Ayat tersebut berbunyi.

وَاللائِي يَئِسْنَ مِنَ الْمَحِيضِ مِنْ نِسَائِكُمْ إِنِ ارْتَبْتُمْ فَعِدَّتُهُنَّ ثَلاثَةُ أَشْهُرٍ وَاللائِي لَمْ يَحِضْنَ وَأُولاتُ الأَحْمَالِ أَجَلُهُنَّ أَنْ يَضَعْنَ حَمْلَهُنَّ وَمَنْ يَتَّقِ اللَّهَ يَجْعَلْ لَهُ مِنْ أَمْرِهِ يُسْرًا

"Dan perempuan-perempuan yang tidak haid lagi (menopause) di antara perempuan-perempuanmu jika kamu ragu-ragu (tentang masa idahnya) maka didah mereka adalah tiga bulan; dan begitu (pula) perempuan-perempuan yang tidak haid. Dan perempuan-perempuan yang hamil, waktu idah mereka itu ialah sampai mereka melahirkan kandungannya. Dan barang siapa yang bertakwa kepada Allah niscaya Allah menjadikan baginya kemudahan dalam urusannya." (Q.S. At-Thalaq: 4)

Syekh Abu Bakar ibn Muhammad al-Husani dalam kitab *Kifayatul Akhyar* menjelaskan bahwa *iddah* adalah masa tunggu bagi seorang wanita untuk mengetahui kekosongan rahimnya. Kekosongan tersebut dapat diketahui dengan kelahiran, hitungan bulan, atau hitungan quru' (masa suci). Hal ini menunjukkan bahwa masa *iddah* berperan penting untuk memastikan kesehatan reproduksi wanita.

Dalam konteks yang sama, masa *iddah* yang diterapkan pada wanita yang mengalami penceraian hidup atau perceraian mati hal ini mencerminkan perhatian Islam terhadap kesehatan dan kesejahteraan wanita. Masa *iddah* memberikan kesempatan bagi wanita untuk memulihkan diri secara fisik dan emosional, memastikan bahwa mereka tidak hamil, dan memungkinkan pemantauan kesehatan reproduksi yang lebih baik. Dalam memperhatikan sebab dan kondisi, wanita yang

menjalani masa *iddah* secara umum terbagi menjadi enam kondisi: (1) wanita yang ditinggal wafat suami dalam keadaan hamil, (2) wanita yang ditinggal wafat suami tidak dalam keadaan hamil, (3) wanita yang dicerai suami dalam keadaan hamil, (4) wanita yang dicerai suami tidak dalam keadaan hamil, sudah pernah bergaul suami-istri, dan sudah/masih haid, (5) wanita yang dicerai tidak dalam keadaan hamil, sudah pernah bergaul suami-istri, dan belum haid/ sudah berhenti haid (menopouse), (6) wanita yang dicerai namun belum pernah bergaul suami-istri.

Alasan di balik penerapan masa *iddah* dalam islam tidak hanya berkaitan dengan aspek spiritual, namun juga mencakup perhatian terhadap kesehatan fisik dan kesejahteraan emosional wanita. Masa *iddah* ini diperlakukan setelah perceraian atau kematian suami yang memberikan kesempatan bagi wanita untuk memulihkan diri secara fisik maupun emosional. Memastikan bahwa tidak hamil dan memungkinkan pemantauan kesehatan reproduksi yang lebih baik. Seperti halnya uji klasifikasi yang bertujuan untuk mencegah dan mendeteksi kanker serviks sejak dini, yang mana masa *iddah* bertujuan untuk melindungi kesehatan wanita dan mencegah kompolikasi yang mungkin timbul dari pernikahan yang terlalu cepat setelah masa duka. Dengan mematuhi ajaran-ajaran ini, umat Islam dapat mengambil langkah-langkah preventif yang signifikan, menjaga kesucian diri, serta memelihara kesehatan jasmani dan jiwa sesuai dengan prinsip-prinsip Islam.

BAB V
KESIMPULAN DAN SARAN

5.1 Kesimpulan

Peneliti menggunakan teknik SMOTE (*Synthetic Minority Over-sampling Technique*) untuk meningkatkan kelas minoritas sehingga setiap kelas memiliki jumlah data yang seimbang. Dengan SMOTE, sampel dari kelas minoritas diperbanyak secara sintetis untuk mengimbangi jumlah data dengan kelas mayoritas. Dalam proses ini, sampel baru dibuat melalui interpolasi antara sampel minoritas yang ada dan tetangga terdekatnya. Saat SMOTE digunakan, target sampel minoritas diperbesar dengan menghasilkan lebih banyak sampel. Akibatnya, nilai variabel atau fitur dari sampel baru juga di SMOTE agar fitur-fitur tersebut menunjukkan distribusi yang seimbang antar kelas. Hasil uji coba akan diberikan dengan PCA dan tanpa PCA, serta dengan metode SVM dengan hyperparameter random terbaik.

Metode PCA (*Principal Component Analysis*) digunakan untuk mengurangi dimensi variabel dalam dataset. Tujuannya adalah untuk menyederhanakan dan menghilangkan faktor atau indikator yang tidak dominan dan relevan tanpa mengurangi tujuan dan maksud dari data awal. Sebelum PCA, SMOTE digunakan untuk menyeimbangkan jumlah sampel untuk masing-masing kelas; kemudian, PCA digunakan untuk mengurangi dimensi variabel dalam dataset. Karena distribusi awal data positif dan negatif tidak sama, hasil komponen utama PCA berbeda untuk setiap target.

Meskipun berbagai pendekatan digunakan, penting untuk memastikan bahwa model tidak *overfitting* dan dapat berfungsi dengan baik pada data yang belum pernah dilihat sebelumnya dalam penelitian ini. Sebelum melakukan reduksi dimensi dengan PCA, penggunaan SMOTE membantu menyeimbangkan distribusi kelas, tetapi tidak meningkatkan akurasi model secara signifikan. PCA tanpa SMOTE menunjukkan *overfitting* yang tinggi, yang berarti bahwa model tidak akan bekerja dengan baik pada data baru; kombinasi terbaik untuk menghindari overfitting adalah dengan menggunakan SMOTE tanpa PCA, yang menghasilkan akurasi yang lebih stabil daripada PCA. Untuk menjamin kinerja yang baik pada data uji yang baru, optimalisasi model harus mempertimbangkan keseimbangan antara akurasi dan kemampuan generalisasi.

5.2 Saran

Peneliti menyadari terdapat beberapa kekurangan dalam penelitian ini dan perlunya kritik dan saran yang membangun agar meningkatkan penelitian selanjutnya. Hal tersebut meliputi:

1. Menggunakan dataset yang lebih besar dan representatif untuk memvalidasi hasil yang diperoleh.
2. Melakukan optimasi lebih lanjut terhadap *hyperparameter* model SVM untuk meningkatkan performa klasifikasi.
3. Menerapkan teknik lain seperti ensemble learning untuk meningkatkan akurasi dan kestabilan model.
4. Mempertimbangkan penggunaan teknik reduksi dimensi fitur selain PCA untuk kasus ini.

DAFTAR PUSTAKA

Abdurrahman, G. (2023). Klasifikasi Kanker Payudara Menggunakan Algoritma SVM dengan Kernel RBF, Linier, dan Sigmoid. *JUSTIFY : Jurnal Sistem Informasi Ibrahimy*, *2*(1), 74–80. https://doi.org/10.35316/justify.v2i1.3370

Almais, A. T. W., Susilo, A., Naba, A., Sarosa, M., Crysdian, C., Tazi, I., Hariyadi, M. A., Muslim, M. A., Basid, P. M. N. S. A., Arif, Y. M., Purwanto, M. S., Parwatiningtyas, D., Supriyono, & Wicaksono, H. (2023). Principal Component Analysis-Based Data Clustering for Labeling of Level Damage Sector in Post-Natural Disasters. *IEEE Access*, *11*(July), 74590–74601. https://doi.org/10.1109/ACCESS.2023.3275852

Andrian, Steele, Salim, E. S., Bindan, H., Pranoto, E., & Dharma, A. (2020). Analisa Metode Random Forest Tree dan K-Nearest Neighbor dalam Mendeteksi Kanker Serviks. *Jurnal Ilmu Komputer Dan Sistem Informasi*, *3*(2), 97–101. http://ejournal.sisfokomtek.org/index.php/jikom/article/view/73

Anggoro, D. A., & Novitaningrum, D. (2021). Comparison of accuracy level of support vector machine (SVM) and artificial neural network (ANN) algorithms in predicting diabetes mellitus disease. *ICIC Express Letters*, *15*(1), 9–18. https://doi.org/10.24507/icicel.15.01.9

Chollet, F. (2017). MACHINE LEARNING. In *Machine Learning* (Vol. 45, Issue 13). https://books.google.ca/books?id=EoYBngEACAAJ&dq=mitchell+machine+learning+1997&hl=en&sa=X&ved=0ahUKEwiomdqfj8TkAhWGslkKHRCbAtoQ6AEIKjAA

Dwi Astuti, F., & Nova Lenti, F. (2021). Implementasi SMOTE untuk mengatasi Imbalance Class pada Klasifikasi Car Evolution menggunakan K-NN. *Jurnal JUPITER*, *13*(1), 89–98.

Ellyzabeth, S. (2018). Pengaruh Pendidikan Kesehatan Tentang Kanker Servik Terhadap Peningkatan Motivasi Untuk Mencegah Kanker Servik. *Global Health Science*, *3*(1), 7–11. http://jurnal.csdforum.com/index.php/GHS/article/view/229

Field, N., & Gilson, R. (2018). Human Papillomavirus (HPV). *Case Studies in Infection Control*, 101–111. https://doi.org/10.1201/9780203733318-9

Gokulnath, C. B., & Shantharajah, S. P. (2019). An optimized feature selection based on genetic approach and support vector machine for heart disease. *Cluster Computing*, *22*(s6), 14777–14787. https://doi.org/10.1007/s10586-018-2416-4

Harimoorthy, K., & Thangavelu, M. (2021). Multi-disease prediction model using improved SVM-radial bias technique in healthcare monitoring system. *Journal of Ambient Intelligence and Humanized Computing*, *12*(3), 3715–

3723. https://doi.org/10.1007/s12652-019-01652-0

Hoq, M., Uddin, M. N., & Park, S. B. (2021). Vocal feature extraction-based artificial intelligent model for parkinson's disease detection. *Diagnostics*, *11*(6). https://doi.org/10.3390/diagnostics11061076

Ihsani, D. A., Arifin, A., & Fatoni, M. H. (2020). Klasifikasi DNA Microarray Menggunakan Principal Component Analysis (PCA) dan Artificial Neural Network (ANN). *Jurnal Teknik ITS*, *9*(1). https://doi.org/10.12962/j23373539.v9i1.51637

Jolliffe, P. M. (2019). Introduction. *Prisons and Forced Labour in Japan*, 1–17. https://doi.org/10.4324/9781351206358-1

Kasanah, A. N., Muladi, M., & Pujianto, U. (2019). Penerapan Teknik SMOTE untuk Mengatasi Imbalance Class dalam Klasifikasi Objektivitas Berita Online Menggunakan Algoritma KNN. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, *3*(2), 196–201. https://doi.org/10.29207/resti.v3i2.945

Kusumawati, Y., Nugrahaningtyas, R. W., & Rahmawati, E. N. (2016). Pengetahuan, Deteksi Dini dan Vaksinasi HPV sebagai Faktor Pencegah Kanker Serviks di Kabupaten Sukoharjo. *Jurnal Kesehatan Masyarakat*, *11*(2), 204. https://doi.org/10.15294/kemas.v11i2.4208

Pradana, M. G., Saputro, P. H., & Wijaya, D. P. (2022). Komparasi Metode Support Vector Machine Dan Naïve Bayes Dalam Klasifikasi Peluang Penyakit Serangan Jantung. *Indonesian Journal of Business Intelligence (IJUBI)*, *5*(2), 87. https://doi.org/10.21927/ijubi.v5i2.2659

Prasetyo, A., Mulia, D. A., & Octavina, K. K. (2023). *INTELLIGENT SYSTEMS AND APPLICATIONS IN Comparative Study of Heart Failure Prediction Algorithm : Logistic Regression and SVM*. *11*(2), 518–522.

Purnama, D. I. (2019). Analisis Komponen Utama Pada Data Potensi Kecamatan di Kota Palu Sebelum Bencana Gempa Bumi dan Tsunami 28 September 2018. *Jurnal Matematika, Statistika Dan Komputasi*, *16*(1), 25. https://doi.org/10.20956/jmsk.v16i1.6329

Puspitasari, A. M., Ratnawati, D. E., & Widodo, A. W. (2018). Klasifikasi Penyakit Gigi Dan Mulut Menggunakan Metode Support Vector Machine. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, *2*(2), 802–810. http://j-ptiik.ub.ac.id

Rahayuwati, L., Rizal, I. A., Pahria, T., Lukman, M., & Juniarti, N. (2020). Pendidikan Kesehatan tentang Pencegahan Penyakit Kanker dan Menjaga Kualitas Kesehatan. *Media Karya Kesehatan*, *3*(1), 59–69. https://doi.org/10.24198/mkk.v3i1.26629

Santoso, I. B., Adrianto, Y., Sensusiati, A. D., Wulandari, D. P., & Purnama, I. K. E. (2022). *Ensemble Jaringan Neural Konvolusional Dengan Mendukung Mesin Vektor untuk Epilepsi Klasifikasi Berdasarkan Multi-Urutan Gambar*

Resonansi Magnetik.

Sari, N. P., Imam, S., Zain, N. C., Asrianti, A. N., Akmal, N. K., Salmahanifah, S., & Aminah, Z. Y. (2023). Perancangan Desain Kemasan Penyedap Rasa Berbasis Kansei Engineering. *Seminar Nasional Inovasi Vokasi*, *2*(1), 1–11.

Sari, N. P. W. P., & Bura Mare, A. C. (2022). Predictors of Quality of Life of Family Caregiver in A Community Setting: Breast and Cervical Cancer Impacts. *Indonesian Journal of Cancer*, *16*(4), 210. https://doi.org/10.33371/ijoc.v16i4.820

Sirait, D. T. C., Adiwijaya, A., & ... (2019). Analisis Perbandingan Reduksi Dimensi Principal Component Analysis (pca) Dan Partial Least Square (pls) Untuk Deteksi Kanker Menggunakan Data …. *EProceedings …*, *6*(2), 8570–8581. https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/viewFile/9828/9689

Wang, W., & Siau, K. (2019). Artificial intelligence, machine learning, automation, robotics, future of work and future of humanity: A review and research agenda. *Journal of Database Management*, *30*(1), 61–79. https://doi.org/10.4018/JDM.2019010104

Zahras, D., & Rustam, Z. (2018). Cervical Cancer Risk Classification Based on Deep Convolutional Neural Network. *Proceedings of ICAITI 2018 - 1st International Conference on Applied Information Technology and Innovation: Toward A New Paradigm for the Design of Assistive Technology in Smart Home Care*, 149–153. https://doi.org/10.1109/ICAITI.2018.8686767

LAMPIRAN-LAMPIRAN

LAMPIRAN

A. Source Code Sequential Minimum Optimization

```
import numpy as np


class SVM:
    def __init__(self, C=1.0, tol=1e-3, max_iter=100, kernel='linear'):
        self.C = C  # Parameter penalti
        self.tol = tol  # Toleransi untuk kriteria berhenti
        self.max_iter = max_iter  # Jumlah iterasi maksimum
        self.kernel = kernel  # Jenis kernel (linear, polynomial, dll.)
        self.alphas = None  # Variabel Lagrange multipliers
        self.b = 0  # Variabel bias
        self.support_vectors = None  # Vektor-vektor pendukung
        self.w = None  # Koefisien w untuk hiperplane


    def fit(self, X, y):
        n_samples, n_features = X.shape
        self.alphas = np.zeros(n_samples)


        for _ in range(self.max_iter):
            # Simpan alpha pada awal iterasi untuk memeriksa konvergensi
            prev_alphas = np.copy(self.alphas)
```

```
for i in range(n_samples):
                # Pilih pasangan alpha_i dan alpha_j
secara acak
                j = self._select_random_index(i,
n_samples)

                # Hitung batasan untuk alpha_j yang
baru
                bounds = self._compute_bounds(i, j, y)

                # Perbarui alpha_i dan alpha_j jika
memenuhi kriteria
                if bounds[0] < bounds[1]:
                    self._update_alphas(i, j, bounds,
y, X)

            # Periksa konvergensi
            if np.linalg.norm(self.alphas -
prev_alphas) < self.tol:
                break

        # Set vektor-vektor pendukung dan nilai bias
        self.support_vectors = X[self.alphas > 0]
        self.b = self._compute_bias(X, y)
        self.w = self._compute_w(X, y)

        # Menghitung nilai 1/2 ||w||^2
        self.objective_value =
self._compute_objective_value()
```

```
        # Periksa kendala
        self.constraints_satisfied =
self.check_constraints(X, y)

    def predict(self, X):
        return np.sign(self.decision_function(X))

    def decision_function(self, X):
        return np.dot(X, self.w) + self.b

    def _compute_kernel_matrix(self, X):
        n_samples = X.shape[0]
        kernel_matrix = np.zeros((n_samples,
n_samples))
        for i in range(n_samples):
            for j in range(n_samples):
                kernel_matrix[i, j] =
self._kernel_function(X[i], X[j])
        return kernel_matrix

    def _kernel_function(self, x1, x2):
        # Untuk sederhana, gunakan kernel linear
        return np.dot(x1, x2)

    def _select_random_index(self, i, n_samples):
        j = i
        while j == i:
            j = np.random.randint(0, n_samples)
        return j

    def _compute_bounds(self, i, j, y):
```

```
        if y[i] != y[j]:
            L = max(0, self.alphas[j] - self.alphas[i])
            H = min(self.C, self.C + self.alphas[j] -
self.alphas[i])
        else:
            L = max(0, self.alphas[i] + self.alphas[j]
- self.C)
            H = min(self.C, self.alphas[i] +
self.alphas[j])
        return L, H


    def _update_alphas(self, i, j, bounds, y, X):
        # Hitung nilai kernel antara sampel i dan j
        kernel_ij = self._kernel_function(X[i], X[i]) +
self._kernel_function(X[j], X[j]) - 2 *
self._kernel_function(X[i], X[j])


        # Hitung gradien
        eta = 2 * kernel_ij -
self._kernel_function(X[i], X[i]) -
self._kernel_function(X[j], X[j])
        if eta >= 0:
            return


        # Perbarui alpha_j
        new_alpha_j = self.alphas[j] - y[j] *
(self._predict_sample(j, X, y) - y[j] * self.b) / eta
        new_alpha_j = min(bounds[1], max(bounds[0],
new_alpha_j))


        # Perbarui alpha_i
        new_alpha_i = self.alphas[i] + y[i] * y[j] *
(self.alphas[j] - new_alpha_j)
```

```
        # Perbarui alpha_i dan alpha_j
        self.alphas[i] = new_alpha_i
        self.alphas[j] = new_alpha_j

    def _predict_sample(self, i, X, y):
        return np.sum(self.alphas *
self._kernel_function(X, X[i, :]))

    def _compute_bias(self, X, y):
        # Hitung nilai bias menggunakan vektor-vektor
pendukung
        bias = 0
        for i, alpha in enumerate(self.alphas):
            if alpha > 0:
                bias += y[i]
                bias -= np.sum(self.alphas * y *
self._kernel_function(X, X[i, :]))
        return bias / np.sum(self.alphas > 0)

    def _compute_w(self, X, y):
        # Menghitung w dari alpha, y dan X
        return np.sum((self.alphas * y)[:, np.newaxis]
* X, axis=0)

    def _compute_objective_value(self):
        # Menghitung nilai 1/2 ||w||^2
        return 0.5 * np.dot(self.w, self.w)

    def check_constraints(self, X, y):
        # Memeriksa apakah y_i (x_i w + b) - 1 ≥ 0
untuk semua i
```

```
        for i in range(len(X)):
            if y[i] * (np.dot(X[i], self.w) + self.b) -
1 < 0:
                return False
        return True
```

B. Source Code Principal Component Analysis

```
class PCA():

  def transform(self, X, n_components):


  covariance_matrix=calculate_covariance_matrix(X)



  eigenvalues,eigenvectors=np.linalg.eig(covariance_mat
  rix)

          idx = eigenvalues.argsort()[::-1]
          eigenvalues = eigenvalues[idx][:n_components]
          eigenvectors  =  np.atleast_1d(eigenvectors[:,
  idx])[:, :n_components]


          X_transformed = X.dot(eigenvectors)

          return X_transformed
```

C. Source Code SMOTE

```
class SMOTE:


    def __init__(self,
                 ratio=100,
                 k_neighbors=6,
                 random_state=None):
```

```
        # check input arguments
        if ratio > 0 and ratio < 100:
            self.ratio = ratio
        elif ratio >= 100:
            if ratio % 100 == 0:
                self.ratio = ratio
            else:
                raise ValueError(
                    'ratio over 100 should be multiples
of 100')
        else:
            raise ValueError(
                'ratio should be greater than 0')

        if type(k_neighbors) == int:
            if k_neighbors > 0:
                self.k_neighbors = k_neighbors
            else:
                raise ValueError(
                    'k_neighbors should be integer
greater than 0')
        else:
            raise TypeError(
                'Expect integer for k_neighbors')

        if type(random_state) == int:
            np.random.seed(random_state)

    def _randomize(self, samples, ratio):
        length = samples.shape[0]
```

```
        target_size = length * ratio
        idx = np.random.randint(length,
size=target_size)

        return samples[idx, :]

    def _populate(self, idx, nnarray):
        for i in range(self.N):
            nn = np.random.randint(low=0,
high=self.k_neighbors)
            for attr in range(self.numattrs):
                dif = (self.samples[nnarray[nn]][attr]
                       - self.samples[idx][attr])
                gap = np.random.uniform()
                self.synthetic[self.newidx][attr] =
(self.samples[idx][attr]
                                                      +
gap * dif)
            self.newidx += 1

    def oversample(self, samples, merge=False):

        if type(samples) == list:
            self.samples = np.array(samples)
        elif type(samples) == np.ndarray:
            self.samples = samples
        else:
            raise TypeError(
                'Expect a built-in list or an ndarray
for samples')

        self.numattrs = self.samples.shape[1]
```

```
        if self.ratio < 100:
            ratio = ratio / 100.0
            self.samples =
self._randomize(self.samples, ratio)
            self.ratio = 100


        self.N = int(self.ratio / 100)
        new_shape = (self.samples.shape[0] * self.N,
self.samples.shape[1])
        self.synthetic = np.empty(shape=new_shape)
        self.newidx = 0


        self.nbrs =
NearestNeighbors(n_neighbors=self.k_neighbors)
        self.nbrs.fit(samples)
        self.knn = self.nbrs.kneighbors()[1]


        for idx in range(self.samples.shape[0]):
            nnarray = self.knn[idx]
            self._populate(idx, nnarray)


        if merge:
            return np.concatenate((self.samples,
self.synthetic))
        else:
            return self.synthetic
```