

BAB III

METODOLOGI PENELITIAN

3.1 Gambaran Umum Objek Penelitian

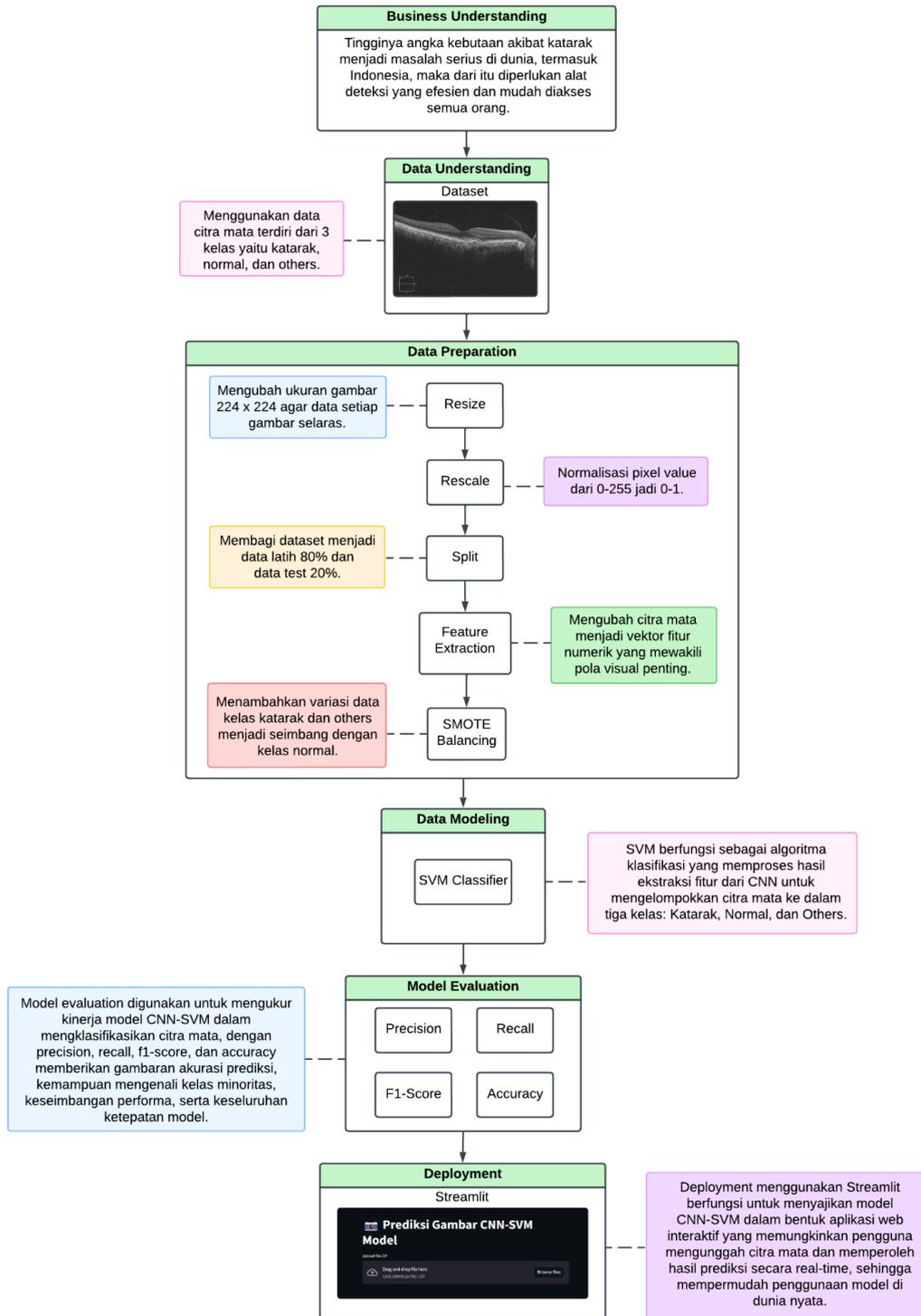
Objek penelitian ini adalah citra mata, yang diperoleh dari Borealis data *Optical Coherence Tomography Image Retinal Database*. Citra mata yang digunakan dalam penelitian ini mencakup data citra mata penderita katarak dan juga mencakup citra mata normal sebagai pembanding dalam proses klasifikasi. Pemilihan dataset yang beragam ini bertujuan untuk memastikan bahwa model *hybrid* CNN-SVM mampu beradaptasi dengan berbagai karakteristik citra dan mendeteksi pola katarak secara efektif. Dalam proses analisis, citra mata akan melalui tahap *pre-processing*, seperti *resize*, *oversampling*, dan augmentasi, untuk meningkatkan kualitas fitur yang akan diekstraksi. Selanjutnya, *feature extraction* akan dilakukan lewat CNN. Dengan menggunakan klasifikasi *Support Vector Machine* (SVM), penelitian ini bertujuan untuk meningkatkan akurasi klasifikasi citra mata dibandingkan dengan metode konvensional, sehingga dapat memberikan solusi diagnosis berbasis kecerdasan buatan yang lebih andal dan presisi.

3.2 Metode Penelitian

3.2.1 Alur Penelitian

Penelitian ini menggunakan alur penelitian CRISP-DM yang digambarkan lewat diagram *flowchart*:

U N I V E R S I T A S
M U L T I M E D I A
N U S A N T A R A



Gambar 3.1 Diagram *Flowchart*

M U L T I M E D I A
N U S A N T A R A

Pada gambar 3.1 merupakan gambaran diagram *flowchart* penelitian ini yang menggunakan metode CRISP-DM yang merupakan salah satu metode *data mining* dan analisis data [17]. Tahapan ini terbagi menjadi 6 tahap utama dalam kerangka kerja ini:

1. *Business Understanding*: Menentukan tujuan bisnis dan kebutuhan untuk memahami data secara lebih menyeluruh. Pada penelitian ini tujuan telah ditentukan yaitu menetapkan 3 masalah dan 3 tujuan sehingga penelitian ini dapat dilakukan.
2. *Data Understanding*: Pengumpulan dan eksplorasi data yang relevan. Pada penelitian ini akan diambil data citra mata penderita katarak sebanyak 55 gambar, mata normal sebanyak 206 gambar, dan penyakit mata lain sebanyak 102 gambar dari Borealis data.
3. *Data Preparation*: Data akan dipersiapkan dan dimodifikasi sebelum melakukan analisis. Pada tahap ini, peneliti mengumpulkan data citra mata penderita katarak sebanyak 55 gambar, mata normal sebanyak 206, dan penyakit mata lain sebanyak 102 gambar pada Borealis data. Setelah data diperoleh, langkah selanjutnya adalah proses *resize dataset* agar setiap gambar memiliki ukuran yang sama. Selanjutnya, dilakukan proses *rescale* yang berguna untuk memperkaya dataset. Dalam penelitian ini, *dataset* akan dibagi menjadi dua bagian, yaitu 80% untuk data *train set* yang digunakan untuk membangun dan melatih model, serta 20% untuk data *test set* yang akan digunakan untuk mengevaluasi kinerja model. Pembagian ini bertujuan untuk memastikan bahwa model dapat mempelajari gambar dengan baik dari data pelatihan, serta diuji akurasi dan kemampuannya dalam mengklasifikasikan gambar pada tahap pengujian. Selain itu, dilakukan juga proses *oversampling* dengan SMOTE yang berguna untuk membuat

dataset antara mata katarak, mata normal dan others seimbang. Data akan dilakukan *split* dengan persentase 80:20 [41]. CNN digunakan dalam penelitian ini sebagai *feature extraction* sebelum model *hybrid* CNN-SVM dikembangkan, *feature extraction* digunakan untuk mengubah citra mata menjadi representasi numerik yang siap digunakan oleh algoritma klasifikasi.

4. *Data Modeling*: Menciptakan dan mengembangkan model algoritma *hybrid* CNN-SVM. Dalam penelitian ini, SVM berfungsi sebagai algoritma klasifikasi yang memproses hasil *feature extraction* dari CNN. Algoritma ini akan diterapkan pada data citra mata penderita katarak, mata normal, dan mata penderita penyakit lain untuk mengklasifikasikan kondisi mata.
5. *Model Evaluation*: Model *hybrid* CNN-SVM akan dinilai dan diuji. Pada penelitian ini metode untuk mengukur performa model, *confusion matrix* akan diterapkan untuk mengukur *accuracy*, *precision*, dan *recall*.
6. *Deployment*: Tahapan *deployment* merupakan tahapan terakhir dari proses CRISP-DM, pada tahapan ini model yang sudah dikembangkan akan dioperasikan dalam ke dalam aplikasi berbasis *website* yaitu Streamlit.

3.2.2 Metode Pembangunan Aplikasi

Metode pengembangan sistem akan memandangkan model dari algoritma CNN dan SVM.

Tabel 3.1 Perbandingan Algoritma

Indikator	CNN	SVM
Pemrosesan	Pemrosesan berbasis citra, mengekstraksi fitur otomatis melalui konvolusi dan <i>pooling</i> .	Pemrosesan berbasis data vektor, memerlukan fitur input yang telah diekstraksi sebelumnya.
Kemampuan	Unggul dalam mengenali pola	Unggul dalam klasifikasi

Indikator	CNN	SVM
Umum	spasial dan fitur kompleks dari gambar.	dengan margin maksimal, terutama pada data berdimensi tinggi.
Kelebihan	Mampu melakukan <i>feature extraction</i> otomatis dari data gambar tanpa pra-pemrosesan manual.	Efektif untuk <i>dataset</i> kecil dan kompleks, tahan terhadap <i>overfitting</i> dengan regularisasi yang kuat.
Kekurangan	Membutuhkan <i>dataset</i> besar dan komputasi tinggi, rawan <i>overfitting</i> jika data terbatas.	Tidak bisa menangani data mentah seperti gambar secara langsung, performa bergantung pada fitur.

Pada tabel 3.1, penelitian ini menggunakan metode *machine learning* dengan algoritma *hybrid* CNN-SVM merupakan pendekatan yang baik karena dengan menggabungkan CNN dan SVM dalam model *hybrid* CNN-SVM, sistem diagnosis menjadi lebih optimal karena CNN bertugas mengekstraksi fitur visual penting dari citra mata, sementara SVM mengklasifikasikan fitur tersebut secara efisien dengan margin optimal. Pendekatan ini menggabungkan kekuatan CNN dalam mengenali pola spasial kompleks dan kekuatan SVM dalam menghasilkan klasifikasi yang akurat pada dataset berdimensi tinggi namun terbatas jumlahnya. Kombinasi ini sangat ideal untuk tugas diagnosis berbasis citra medis seperti katarak, di mana jumlah data seringkali terbatas namun detail spasial sangat penting.

3.2.3 Metode Data Mining

Tabel 3.2 Perbandingan Metode

Indikator	CRISP-DM	SEMMA	KDD
Proses	<ul style="list-style-type: none"> • <i>Business Understanding</i> • <i>Data Understanding</i> • <i>Data Preparation</i> 	<ul style="list-style-type: none"> • <i>Sample</i> • <i>Explore</i> • <i>Modify</i> • <i>Model</i> • <i>Assess</i> 	<ul style="list-style-type: none"> • <i>Selection</i> • <i>Preprocessing</i> • <i>Transformation</i> • <i>Data Mining</i> • <i>Interpretation/Evaluation</i>

Indikator	CRISP-DM	SEMMA	KDD
	<ul style="list-style-type: none"> • <i>Modelling</i> • <i>Evaluation</i> • <i>Deployment</i> 		
Kelebihan	CRISP-DM memiliki kelebihan dalam memberikan struktur yang jelas dan kesesuaian dengan proyek-proyek data mining.	SEMMA menonjol dalam pendekatannya yang lebih sederhana dan <i>linier</i> , cocok untuk proyek-proyek yang lebih kecil	KDD memiliki kelebihan dalam cakupan proses yang menyeluruh dari awal hingga evaluasi hasil, serta fleksibel digunakan untuk eksplorasi pengetahuan dari basis data dalam skala besar.
Kekurangan	CRISP-DM memiliki kekurangan yaitu dapat terjadi kerumitan serta memakan lebih banyak waktu.	Kekurangan dalam penerapan SEMMA ketika menghadapi proyek-proyek dengan kompleksitas pemrosesan dan analisis data yang lebih vertikal.	KDD kurang memberikan panduan teknis yang spesifik di tiap tahap, sehingga dapat membingungkan bagi pengguna non-teknis serta belum menekankan orientasi bisnis seperti pada CRISP-DM.

Pada tabel 3.2, penelitian ini menggunakan metode CRISP-DM sebagai *data mining*. CRISP-DM dipilih karena memiliki kelebihan dalam memberikan struktur yang jelas dan kesesuaian dengan proyek-proyek *data mining*. Contohnya adalah struktur CRISP-DM memiliki kerangka kerja yang jelas yang terdiri dari *Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation, dan Deployment*. Tidak hanya itu, CRISP-DM juga berfokus pada pemahaman bisnis yang berarti ini menegaskan betapa pentingnya melibatkan aspek-aspek bisnis dalam tahap awal proses analisis data, hal ini bertujuan untuk menciptakan solusi akhir yang tepat serta efektif guna mencapai kesesuaian antara hasil penelitian dengan tujuan yang ingin dicapai oleh sebuah organisasi atau perusahaan secara umum. Tetapi CRISP-DM juga memiliki

kekurangan yaitu dapat terjadi kerumitan serta memakan lebih banyak waktu, jika diperlukan iterasi atau revisi pada tahap tertentu maka harus kembali ke awal, maka hal ini bisa menjadi suatu kendala bagi metodologi CRISP-DM.

3.3 Teknik Pengumpulan Data

Pada penelitian ini teknik yang digunakan dalam mengumpulkan data dengan mengambil data dari Borealis data. Data yang didapatkan berasal dari University of Waterloo Dataverse Collection yang berisikan *Optical Coherence Tomography Image Retinal Database*. Data yang didapatkan berisi citra mata penderita katarak sebanyak 55 gambar, mata normal sebanyak 206 gambar, dan penyakit mata lain sebanyak 102 gambar.

3.4 Teknik Analisis Data

Tabel 3.3 Perbandingan *Tools*

Sumber: [37]

Indikator	Google Colab	Spyder
Lingkungan Kerja	Berbasis <i>cloud</i> , dapat diakses langsung melalui browser tanpa instalasi tambahan.	Berbasis <i>desktop</i> , memerlukan instalasi di perangkat lokal.
Integrasi <i>Library</i>	Mendukung hampir semua <i>library</i> Python, dengan dukungan integrasi otomatis.	Memiliki integrasi <i>library</i> yang baik, tetapi pengguna perlu mengatur manual.
Pemrosesan Gambar	Mendukung GPU/TPU untuk pemrosesan data intensif, termasuk pengolahan gambar.	Mengandalkan CPU perangkat lokal, performa bergantung pada spesifikasi sistem.
Kompatibilitas Ekstensi	Mendukung ekstensi <i>notebook</i> seperti Jupyter dengan berbagai fitur tambahan.	Terintegrasi dengan editor Python, tetapi kurang fleksibel dibandingkan <i>notebook</i> .

Pada penelitian ini menggunakan Google Colab, berdasarkan tabel 3.3 di atas, Google Colab menawarkan lingkungan kerja berbasis *cloud* yang dapat

diakses langsung melalui browser tanpa perlu instalasi, menjadikannya lebih fleksibel untuk kolaborasi dan pekerjaan jarak jauh. Selain itu, Colab mendukung hampir semua *library* Python dengan integrasi otomatis dan menyediakan dukungan GPU/TPU, yang sangat membantu dalam pemrosesan data besar seperti pengolahan gambar. Di sisi lain, Spyder merupakan lingkungan kerja berbasis *desktop* yang harus diinstal pada perangkat lokal, dengan integrasi *library* yang memerlukan konfigurasi manual. Spyder mengandalkan CPU untuk pemrosesan data, sehingga performanya bergantung pada spesifikasi perangkat yang digunakan. Meskipun Spyder terintegrasi dengan *editor* Python, fleksibilitasnya dalam kompatibilitas ekstensi masih kalah dibandingkan dengan Google Colab, terutama dalam mendukung fitur *notebook* interaktif.

3.4.1 Variabel Penelitian

3.4.1.1 Variabel Independen

Variabel independen adalah jenis variabel yang sengaja diubah atau dimodifikasi dalam penelitian untuk mengetahui pengaruhnya terhadap variabel dependennya. Dalam penelitian ini, variabel independen nya adalah gambar-gambar dari mata penderita katarak sebanyak 55 gambar, mata normal sebanyak 206 gambar, dan penyakit mata lain sebanyak 102 gambar.

3.4.1.2 Variabel Dependen

Variabel dependen digunakan untuk mengamati dan mengukur hasil dari variasi dalam variabel independennya. Dalam penelitian ini, variabel dependen nya adalah hasil model citra mata lewat algoritma *hybrid* CNN-SVM. Variabel juga diukur dengan menggunakan *confusion matrix* untuk menguji akurasi dari model algoritma yang digunakan.