

BAB III METODE PENELITIAN

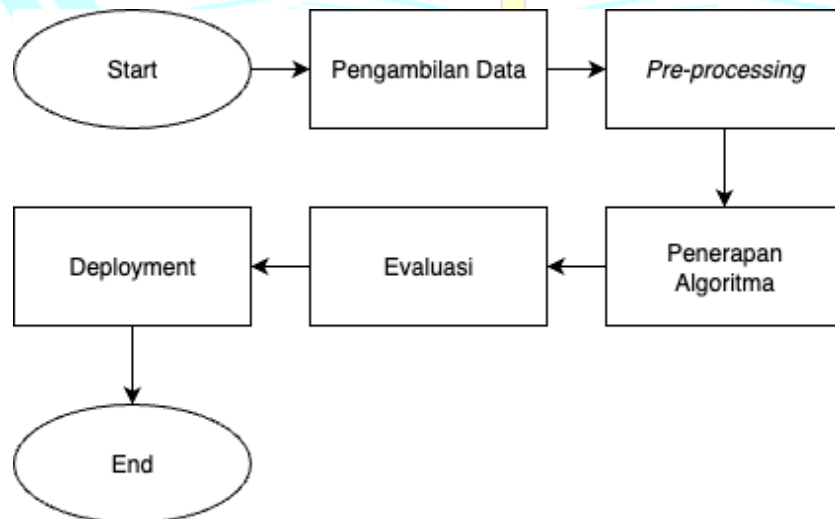
3.1 Objek Penelitian

Penelitian ini berfokus pada penerapan dan perbandingan algoritma *machine learning* dalam mendeteksi penyakit diabetes, khususnya menggunakan Support Vector Machine (SVM) dan *Multi-Layer Perceptron* (MLP). Objek yang dikaji meliputi efektivitas dan akurasi model yang dihasilkan oleh kedua algoritma berdasarkan dataset *Pima Indians Diabetes* dari *UCI Machine Learning Repository*.

Kinerja model dievaluasi menggunakan Confusion Matrix yang mencakup empat elemen, yaitu True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN). Berdasarkan nilai elemen Confusion Matrix, metrik evaluasi seperti *accuracy*, *precision*, *recall*, dan *F1-score* dihitung untuk menunjukkan tingkat efektivitas algoritma SVM dan MLP dalam mendeteksi penyakit diabetes.

3.2 Prosedur Penelitian

Alur yang dilakukan dalam penelitian ini digambarkan sebagai berikut:



Gambar 1.1 Alur penelitian

3.2.1 Pengambilan Data

Tahap ini bertujuan untuk mengambil data dari sumber data dan mempelajarinya guna mengetahui kondisi dan karakteristik dari sebuah data. merupakan tahap pengambilan dataset yang akan digunakan dalam penelitian ini. Penelitian ini menggunakan dataset yang diterbitkan oleh UCI Machine Learning

bernama Pima Indian Diabetes Dataset yang diperoleh dari sumber website *Kaggle* (<https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database>). Gambar 1.2 menunjukkan contoh dari isi dataset.

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Outcome	Age
0	6	148	72	35	0	33.6	0.627	1	50
1	1	85	66	29	0	26.6	0.351	0	31
2	8	183	64	0	0	23.3	0.672	1	32
3	1	89	66	23	94	28.1	0.167	0	21
4	0	137	40	35	168	43.1	2.288	1	33
...
763	10	101	76	48	180	32.9	0.171	0	63
764	2	122	70	27	0	36.8	0.340	0	27
765	5	121	72	23	112	26.2	0.245	0	30
766	1	126	60	0	0	30.1	0.349	1	47
767	1	93	70	31	0	30.4	0.315	0	23

Gambar 1.2 Contoh isi dataset

3.2.2 Pre-processing

Tahap awal *pre-processing* dalam penelitian ini melibatkan pembersihan data (*data cleaning*) melalui penghapusan *missing values* dan eliminasi *outliers*. Metode yang digunakan untuk mendeteksi *outlier* menggunakan *Interquartile Range* (IQR). Pada tahap ini, data dengan nilai yang hilang (*missing values*) akan dihapus terlebih dahulu, kemudian dilanjutkan dengan proses eliminasi *outlier*. Selanjutnya, kondisi keseimbangan distribusi data dianalisis untuk menangani *Class Imbalance* berdasarkan kelas sasaran. Dalam hal ini, atribut *Outcome* digunakan sebagai kelas data. Penelitian ini menangani *Class Imbalance* pada data dengan menerapkan metode SMOTE-ENN. Penelitian ini juga melakukan normalisasi data (*data normalization*) dengan menerapkan teknik *Z-score normalization* untuk meminimalkan perbedaan jarak yang signifikan antar fitur yang bersangkutan dalam dataset guna mengurangi kesalahan dalam analisis data dan meningkatkan kinerja model *machine learning*. Selanjutnya, dataset dibagi menjadi beberapa jenis data yaitu data latih, data validation, dan data uji dengan rasio tertentu.

3.2.3 Penerapan Algoritma

Tahap ini merupakan fase penerapan algoritma dan pembangunan model *machine learning* dengan menggunakan dataset *Pima Indians Diabetes* yang telah melalui proses *pre-processing* pada tahap sebelumnya. Mengingat bahwa algoritma

Support Vector Machine (SVM) dan *Multi-Layer Perceptron* (MLP) memiliki karakteristik dan kebutuhan input data yang berbeda, maka pada tahap ini dilakukan pemilihan jenis data dan transformasi data yang disesuaikan dengan kebutuhan masing-masing algoritma. Proses ini mencakup pengaturan parameter, pembuatan arsitektur, serta penerapan teknik *encoding* yang diperlukan oleh algoritma. Setelah data dipersiapkan, model SVM dan MLP kemudian dilatih menggunakan dataset. Langkah-langkah dari algoritma *Support Vector Machine* dan *Multi-Layer Perceptron* dijelaskan pada Algoritma 1 dan Algoritma 2.

Algoritma 1 Support Vector Machine

Algoritma *Support Vector Machine*

Input: Data latih

Output: Model SVM

1. Pemilihan kernel
 2. Penentuan parameter
 3. Melakukan pemetaan data menjadi representasi vektor berdimensi tinggi
 4. Melakukan pencarian *hyperplane* optimal dengan memaksimalkan margin berdasarkan persamaan (4)
 5. Memperbaharui nilai bias dan nilai *Lagrange Multiplier* berdasarkan persamaan (8)
 6. Ulangi langkah 3-4 sampai memenuhi batas (*threshold*) konvergensi
-

Algoritma 2 Multi-Layer Perceptron

Algoritma *Multi-Layer Perceptron*

Input: Data latih dan data validasi

Output: Model MLP

1. Persiapan data
 2. Persiapan arsitektur MLP.
 3. Inisialisasi bobot neuron dengan nilai acak.
 4. Melakukan *forward propagation* dengan menghitung output dari setiap neuron berdasarkan input data menggunakan persamaan (10)
 5. Menghitung nilai *loss* menggunakan *loss function* untuk mengukur *error* dari prediksi (hasil *forward propagation*).
-

-
6. Melakukan *backward propagation* untuk mengkoreksi bobot neuron dalam masing-masing lapisan berdasarkan *learning rate*.
 7. Melakukan validasi model dengan menggunakan data validasi.
 8. Ulangi langkah 3-5 sampai nilai *loss* mendekati 0 (nol) atau proses pelatihan telah diulang sebanyak N kali.
-

3.2.4 Evaluasi

Salah satu tujuan penelitian ini adalah untuk mengetahui seberapa efektif SVM dan MLP dalam mendeteksi penyakit diabetes. Tahap ini merupakan tahap yang bertujuan untuk mengevaluasi dan menilai kinerja masing-masing model yang dihasilkan oleh algoritma berdasarkan metrik evaluasi. Penelitian ini menggunakan metrik evaluasi seperti *Confusion Matrix*, *accuracy*, *precision*, *recall*, dan *f1-score* untuk menentukan tingkat efektivitas algoritma SVM dan MLP dalam mendeteksi penyakit diabetes berdasarkan data *Pima Indians Diabetes*.

3.2.5 Deployment

Pada tahap ini dilakukan pembuatan sebuah aplikasi sederhana yang mengintegrasikan model hasil penelitian. Aplikasi ini dirancang sebagai *proof of concept* untuk menunjukkan bagaimana model dapat diterapkan. Selain berfungsi sebagai contoh implementasi praktis, aplikasi ini juga diharapkan menjadi dasar yang dapat dikembangkan lebih lanjut.

Aplikasi yang dibuat merupakan aplikasi inferensi penyakit diabetes. Pengguna aplikasi dapat memasukkan parameter-parameter terkait dan sistem akan menentukan apakah seseorang positif atau negatif diabetes. Aplikasi dirancang menggunakan bahasa pemrograman *Python* dan *JavaScript* dengan beberapa teknologi pendukung yang digunakan untuk membantu perancangan aplikasi, antara lain *Flask web framework*, dan pustaka antarmuka *Solid.js*