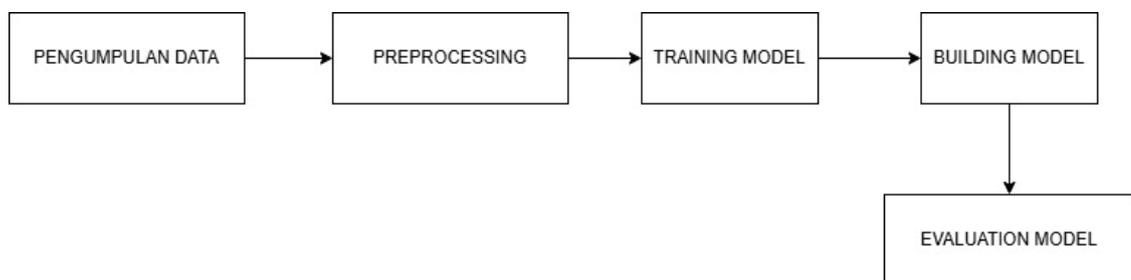


BAB III

METODELOGI PENELITIAN

Metode penelitian ini meliputi beberapa tahapan yang dapat disimpulkan melalui alur metode berikut.



Gambar 3.1 Alur Metode (Image by Author)

Pada flowchart di atas menjelaskan beberapa alur pengembangan yang digunakan untuk klasifikasi kualitas udara yang ada di DKI Jakarta, mulai dari pengumpulan data, preprocessing, training model, building model, dan evaluation model untuk memastikan bahwa sistem menghasilkan nilai yang akurat.

3.1 Metode Pengumpulan Data

Pada tahap pengumpulan data, beberapa teknik digunakan untuk mendapatkan informasi tentang tujuan penelitian ini, seperti:

a. Data Sekunder

Mengakses dan mengunduh dataset publik yang berkaitan dengan kualitas udara dari sumber terpercaya. Dataset dapat diperoleh dari badan pemerintah, seperti Kementerian Lingkungan Hidup, atau organisasi internasional seperti World Health Organization (WHO). Data yang diunduh harus mencakup variabel penting seperti konsentrasi polutan (misalnya PM10, PM2.5, CO2, dan NOx), lokasi pengukuran, serta tanggal dan waktu pengambilan data.

b. Studi Pustaka

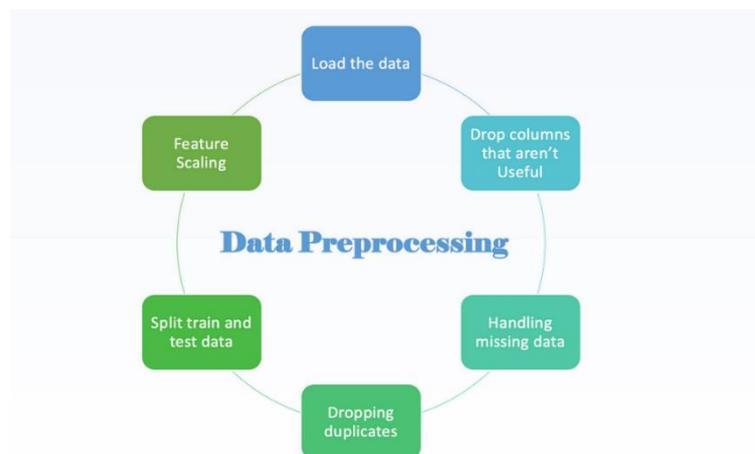
Melakukan kajian literatur untuk mengidentifikasi penelitian sebelumnya dan sumber informasi terkait kualitas udara. Literatur yang ditinjau meliputi jurnal, artikel ilmiah, dan laporan dari organisasi lingkungan yang berfokus pada kualitas udara. Penelitian ini akan memberikan landasan teori dan konteks yang diperlukan untuk pemahaman lebih mendalam.

c. Pengumpulan Data

Pengumpulan data yang dilakukan untuk melakukan pemantauan kualitas udara yang ada di Jakarta dilakukan dengan mengambil sumber data yang berasal dari dataset satu data Jakarta, yang akan diolah melalui proses train data dalam penelitian yang sedang dilakukan.

3.1.1 Pre-processing

Sebelum melakukan sebuah training data, maka akan dilakukan pre-processing data agar dataset yang akan digunakan sesuai dengan model. Selanjut nya akan dilakukan normalisasi yang dimana PM2.5 memiliki nilai 0-500, sementara suhu berkisar 10-40. Akan di normalisasi memastikan semua fitur berada dalam rentang [0,1]. Hal ini bertujuan untuk mempermudah proses training data dan juga membantu model dalam mencapai konvergensi dengan lebih efisien.



Gambar 3.2 Preprocessing Data

Sumber : medium.com, Data Preprocessing in Python

3.1.2 Building Model

Pada tahap ini, Proses pembentukan dan pelatihan model **Support Vector Machine (SVM)** dilakukan untuk mengklasifikasikan kualitas udara berdasarkan parameter yang telah ditentukan. SVM dipilih karena kemampuannya dalam mengolah data dengan dimensi tinggi serta keunggulannya dalam menentukan hyperplane optimal yang memisahkan kategori dengan margin terbesar.

Proses ini dimulai dengan memilih kernel function yang paling sesuai untuk menangani kompleksitas data, seperti Linear, Polynomial, Radial Basis Function (RBF), atau Sigmoid, agar model dapat bekerja secara optimal. Setelah itu, dilakukan penyesuaian hyperparameter menggunakan metode Grid Search atau Random Search untuk menemukan nilai terbaik dari C (regularization parameter) dan γ (gamma parameter untuk RBF kernel). Hal ini bertujuan agar model dapat mencapai keseimbangan yang baik antara bias dan varians, sehingga mampu memberikan prediksi yang lebih akurat dan stabil.

3.1.3 Training Model

Pada tahap ini, model **Support Vector Machine (SVM)** dilatih menggunakan dataset yang telah diproses sebelumnya. Data dibagi menjadi **training set** dan **validation set** agar model dapat belajar pola dengan baik tanpa mengalami **overfitting**.

Selama pelatihan ini, SVM mencari hyperplane terbaik yang dapat memisahkan data dengan margin maksimal. Proses ini menggunakan algoritma Sequential Minimal Optimization (SMO) yang dimana dapat mempercepat pencarian solusi. Selain itu, dilakukan penyesuaian hyperparameter, seperti C (regularization) dan γ (gamma), dengan teknik Grid Search atau Random Search, sehingga model dapat menemukan keseimbangan yang tepat antara akurasi dan kompleksitas.

3.1.4 Evaluation Model

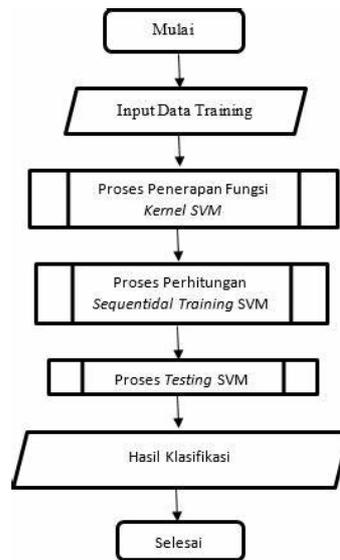
Setelah model **Support Vector Machine (SVM)** selesai dilatih, langkah selanjutnya adalah mengevaluasi performanya untuk memastikan bahwa model dapat membuat prediksi dengan akurat dan andal. Proses ini dilakukan dengan menguji model menggunakan **testing set**, yaitu data yang belum pernah digunakan dalam tahap training, sehingga dapat menunjukkan sejauh mana model mampu mengenali pola baru di dunia nyata.

Beberapa metrik utama digunakan untuk menilai kinerja model, seperti akurasi untuk melihat seberapa sering prediksi model benar, precision dan recall untuk mengukur ketepatan serta sensitivitas model terhadap data, serta F1-score yang memberikan gambaran keseimbangan antara precision dan recall. Selain itu, confusion matrix digunakan untuk memahami kesalahan prediksi, sementara ROC Curve dan AUC membantu menilai seberapa baik model membedakan antara kategori yang berbeda.

Jika hasil evaluasi menunjukkan bahwa model masih belum optimal, maka dilakukan fine-tuning hyperparameter atau penyesuaian lebih lanjut agar model dapat bekerja lebih efisien. Dengan evaluasi yang matang, model ini diharapkan dapat memberikan prediksi kualitas udara yang akurat dan dapat diandalkan dalam penerapan nyata.

3.2 Algoritma Support Vector Machine (SVM)

Pemilihan algoritma Support Vector Machine (SVM) didasarkan pada efektivitasnya dalam melakukan klasifikasi data, terutama ketika data memiliki dimensi yang tinggi. Dalam konteks pemantauan kualitas udara, SVM dapat digunakan untuk mengklasifikasikan tingkat polusi atau kualitas udara berdasarkan berbagai variabel input, seperti konsentrasi polutan. Alur proses dari algoritma SVM ditunjukkan pada Gambar 3.3.



Gambar 3.3 Alur Proses Model SVM

3.3 Dataset

Pengumpulan dataset dilakukan melalui bank data portal <https://satudata.jakarta.go.id/> yang dimana memiliki banyak data berdasarkan bulan. Penelitian ini menggunakan data dari bulan Februari hingga November 2023, mencakup periode Februari, Maret, April, Mei, Juni, Juli, Agustus, September, Oktober dan November 2023. Data ini bersumber dari <https://data.go.id/dataset/dataset/data-indeks-standar-pencemar-udara-ispu-di-provinsi-dki-jakarta-2023>, dengan total 1826 catatan yang telah terdokumentasi.

Dalam proses pengolahan data, dilakukan penyaringan dan pengurangan jumlah atribut agar informasi yang digunakan tetap relevan dan tidak berlebihan. Beberapa variabel utama yang digunakan sebagai input meliputi tanggal, lokasi, PM10, PM2.5, SO₂, CO, O₃, dan NO₂, karena faktor-faktor ini memiliki pengaruh besar terhadap kualitas udara. Sementara itu, variabel output yang digunakan adalah kategori kualitas udara, yang nantinya akan dianalisis lebih lanjut dalam tahap pre-processing. Langkah awal yang perlu dilakukan adalah mengolah dataset sesuai dengan yang ditunjukkan pada gambar 3.2.

periode_data	tanggal	stasiun	pm_sepuluh	pm_duakomalima	sulfur_dioksida	karbon_monoksida	ozon	nitrogen_dioksida	max	parameter_pencemar_kritis	kategori
0	202302	2023-02-25 DKI5 Kebon Jeruk Jakarta Barat	35	-	13	12	31	18	35	PM10	BAIK
1	202302	2023-02-26 DKI5 Kebon Jeruk Jakarta Barat	23	-	14	9	32	11	32	03	BAIK
2	202302	2023-02-27 DKI5 Kebon Jeruk Jakarta Barat	20	-	13	8	33	13	33	03	BAIK
3	202302	2023-02-28 DKI5 Kebon Jeruk Jakarta Barat	30	-	21	11	28	18	30	PM10	BAIK
4	202303	2023-03-01 DKI1 Bunderan HI	38	44	50	8	19	27	50	3	BAIK
...
1820	202311	2023-11-26 DKI5 Kebon Jeruk	48	71	33	21	44	20	71	PM25	SEDANG
1821	202311	2023-11-27 DKI5 Kebon Jeruk	51	76	32	18	53	19	76	PM25	SEDANG
1822	202311	2023-11-28 DKI5 Kebon Jeruk	54	80	32	20	48	21	80	PM25	SEDANG

Gambar 3.4 Dataset Satu Data DKI JAKARTA