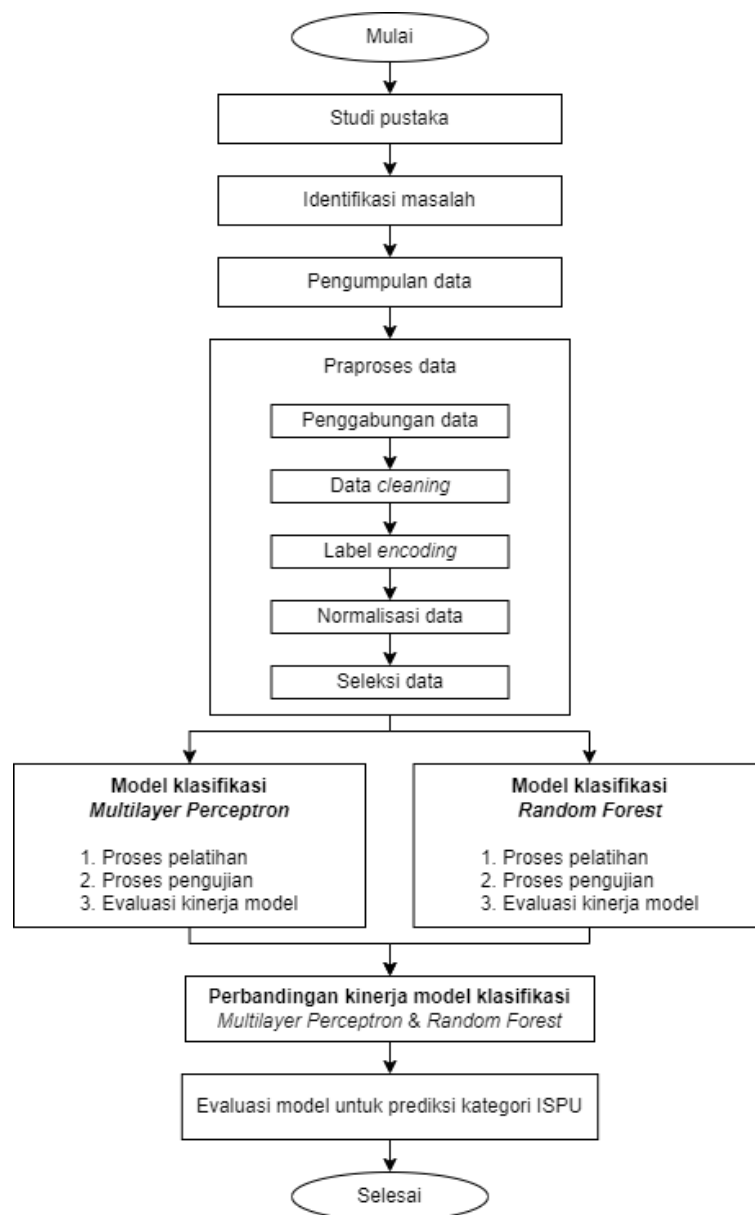


BAB III

METODOLOGI

Tahapan yang dilalui pada penelitian klasifikasi kategori Indeks Standar Pencemar Udara (ISPU) di DKI Jakarta dengan menerapkan metode *Multilayer Perceptron* dan *Random Forest* digambarkan pada Gambar 3.1.



Gambar 3.1 Tahapan Penelitian

3.1 Studi Pustaka

Studi pustaka dilakukan untuk mengumpulkan informasi yang relevan dengan topik yang menjadi objek penelitian. Sumber studi pustaka yang dilakukan berupa buku, *e-book*, jurnal nasional dan beberapa jurnal internasional.

Informasi yang dikumpulkan berkaitan dengan kualitas udara, Indeks Standar Pencemar Udara (ISPU), *Machine Learning*, metode *Multilayer Perceptron*, dan metode *Random Forest*.

3.2 Identifikasi Masalah

Identifikasi masalah merupakan proses mendefinisikan masalah serta membuat definisi tersebut menjadi lebih terukur dan dianggap bisa ditemukan jawabannya melalui sebuah penelitian yang dilakukan secara ilmiah.

Dalam proses penerapan metode *Multilayer Perceptron* (MLP) dan metode *Random Forest* untuk kasus klasifikasi diperlukan data yang berkaitan dengan ISPU. Untuk mendapat kinerja metode MLP dan *Random Forest* yang baik, diperlukan nilai optimal yang diperoleh dari konfigurasi parameter kedua metode.

3.3 Pengumpulan Data

Data Indeks Standar Pencemaran Udara (ISPU) di DKI Jakarta tahun 2021 tersedia pada laman website Jakarta Open Data dengan alamat <https://data.jakarta.go.id/>. Dataset berisi mengenai ISPU yang diukur dari 5 stasiun pemantau udara (SPKU) yang ada di provinsi DKI Jakarta selama tahun 2021.

3.4 Praproses Data

Praproses data atau *pre-processing* data dilakukan dengan bahasa pemrograman *Python*. Proses yang dilakukan diantaranya:

1) Penggabungan data

Data ISPU di DKI Jakarta tersedia dalam rentang waktu satu bulan, untuk menggunakan data satu tahun maka dilakukan penggabungan data dari bulan Januari sampai dengan Desember 2021.

2) *Cleaning* data

Cleaning data dilakukan untuk membersihkan data dari komponen yang tidak sesuai, salah satunya adalah data yang tidak memiliki nilai.

3) *Label encoding*

Label encoding dilakukan untuk mengubah tipe data *string* pada variabel kategori menjadi tipe numerik. Pada *python*, *label encoding* mengganti nilai kategorikal dengan nilai numerik antara 0 dan jumlah kelas kurang 1. Jika kategori ISPU terdapat baik, sedang, dan tidak sehat, maka akan diubah kedalam index 0,1, dan 2.

4) Normalisasi data

Normalisasi data dilakukan untuk mengubah variabel numerik kedalam rentang bilangan yang sama. Normalisasi data dengan *min-max normalization* mengubah rentang data menjadi [0, 1]. Untuk setiap fitur, nilai minimum fitur tersebut diubah menjadi 0, nilai maksimum diubah menjadi 1, dan setiap nilai lainnya diubah menjadi desimal antara 0 dan 1.

5) Seleksi data

Dari 11 variabel pada setiap data ISPU yang tersedia, hanya konsentrasi parameter polutan PM₁₀, PM_{2.5}, SO₂, CO, O₃, NO₂, dan kategori ISPU yang digunakan selama penelitian.

3.5 Model Klasifikasi

Klasifikasi kategori ISPU dilakukan dengan dua metode *machine learning* dengan menggunakan bahasa pemrograman *python*, yaitu:

1) *Multilayer Perceptron*

Klasifikasi kategori ISPU dengan metode *Multilayer Perceptron* (MLP) memanfaatkan *library* modul *keras* model *sequential*.

Pada metode MLP, beberapa hyperparameter yang diatur yaitu jumlah node pada setiap *layer*, jumlah *hidden layer*, fungsi aktivasi, *loss fuction*, dan *optimizer*. Evaluasi kinerja model dilakukan dengan perhitungan nilai *accuracy*, *precision*, *recall*, dan *f1-score*.

2) *Random Forest*

Klasifikasi kategori ISPU dengan metode *Random Forest* memanfaatkan *library* modul *sklearn* yaitu *RandomForestClassifier*.

Pada metode *Random Forest*, beberapa parameter yang diatur yaitu *criterion*, *n_estimator*, *bootsrap sampling*, dan nilai *Out-of-Bag*. Evaluasi kinerja model dilakukan dengan perhitungan nilai *accuracy*, *precision*, *recall*, dan *f1-score*.

3.6 Perbandingan Model Klasifikasi

Analisis perbandingan model klasifikasi MLP dan *Random Forest* dilakukan dengan membandingkan kinerja model untuk nilai *accuracy*, *precision*, *recall*, dan *f1-score*. Kemudian menentukan model klasifikasi dengan kinerja paling optimal.

3.7 Evaluasi Model Untuk Prediksi Kategori ISPU

Evaluasi untuk memprediksi kategori ISPU pada data polutan baru yang belum diketahui nilai kategorinya dilakukan dengan metode klasifikasi yang paling optimal setelah dilakukan analisis antara MLP dan *Random Forest*. Evaluasi dilakukan untuk melihat apakah model klasifikasi dapat mengenali pola data polutan dan memprediksi kategori ISPU.