

# BAB 1

## ANALISIS KEBUTUHAN

### 1.1 Latar Belakang Masalah

Menurut data dari IQAir, pada bulan Oktober 2021, Ibu kota Indonesia, Jakarta, menduduki peringkat ke-9 untuk kualitas udara dan polusi kota. Indonesia menempati peringkat ke-9 dari 106 negara untuk negara paling berpolusi di dunia pada tahun 2020 dari konsentrasi PM<sub>2.5</sub>[1]. Maka dari itu, penelitian ini dilakukan untuk memprediksi dan mengklasifikasikan kualitas udara.

DKI Jakarta merupakan kota yang memiliki kepadatan kendaraan bermotor yang tinggi. 70% merupakan polutan yang disebabkan oleh emisi dari kendaraan bermotor. Kendaraan bermotor merupakan penghasil polusi terbesar karena mengeluarkan banyak sekali zat yang sangat berbahaya dan merusak kesehatan manusia termasuk merusak lingkungan juga[2].

Pada penelitian sebelumnya, telah dilakukan prediksi untuk memperkirakan kualitas udara pada hari kedepannya yang divisualisasikan melalui halaman web dengan menggunakan metode prediksi yaitu *Gaussian Process Regression* (GPR). Pada penelitian ini, selain melakukan prediksi, data yang didapatkan akan diklasifikasikan dan akan divisualisasikan melalui website. Metode yang akan digunakan adalah metode *Extreme Learning Machine* (ELM) dan *Kernel Extreme Learning Machine* (K-ELM) untuk proses regresi. Sedangkan untuk klasifikasi data kami gunakan metode *Extreme Learning Machine* (ELM).

Metode *Extreme Learning Machine* (ELM) digunakan karena metode ini memiliki kelebihan pembelajaran lebih cepat, mudah untuk diaplikasikan pada masalah kompleks dan diterapkan pada kehidupan nyata[3]. *Kernel Extreme Learning Machine* (K-ELM) merupakan salah satu evolusi dari metode *Extreme Learning Machine* (ELM). Perbedaan signifikan antara metode *Kernel ELM* dengan *ELM* sebelumnya yakni pada pemanfaatan fungsi kernel di hidden layer-nya. Selain itu, dibandingkan dengan *ELM*, pemetaan fitur hidden layer tidak perlu diketahui dan jumlah hidden neuron juga tidak perlu dipilih. *Kernel ELM* dapat mengatasi masalah variasi pada *ELM* dan meningkatkan stabilitasnya[4]. Prediksi dan indeks yang telah dilakukan akan diimplementasikan kedalam website sebagai hasil visualisasi yang telah didapatkan.

## 1.2 Informasi Pendukung

Data kualitas udara yang digunakan yaitu data Indeks Standar Pencemaran Udara (ISPU) yang didapat dari portal website Jakarta Open Data[5]. Dikarenakan ada beberapa data yang hilang, maka digunakan metode K-Nearest Neighbor (KNN) Imputer untuk melakukan pengisian data yang hilang[6]. Metode *Extreme Learning Machine* (ELM) digunakan karena metode ini memiliki kelebihan pembelajaran lebih cepat, mudah untuk diaplikasikan pada masalah kompleks dan diterapkan pada kehidupan nyata[3]. *Kernel Extreme Learning Machine* (K-ELM) merupakan salah satu evolusi dari metode *Extreme Learning Machine* (ELM). Perbedaan signifikan antara metode Kernel ELM dengan ELM sebelumnya yakni pada pemanfaatan fungsi kernel di hidden layer-nya. Selain itu, dibandingkan dengan ELM, pemetaan fitur hidden layer tidak perlu diketahui dan jumlah hidden neuron juga tidak perlu dipilih. Kernel ELM dapat mengatasi masalah variasi pada ELM dan meningkatkan stabilitasnya[4]. Menurut data dari IQAir, pada bulan Oktober 2021, Ibu kota Indonesia, Jakarta, menduduki peringkat ke-9 untuk kualitas udara dan polusi kota. Indonesia menempati peringkat ke-9 dari 106 negara untuk negara paling berpolusi di dunia pada tahun 2020 dari konsentrasi PM<sub>2.5</sub>[1]. Kendaraan bermotor merupakan penghasil polusi terbesar karena mengeluarkan banyak sekali zat yang sangat berbahaya dan merusak kesehatan manusia termasuk merusak lingkungan juga[2]. Salah satu faktor yang paling berpengaruh terhadap kesehatan manusia adalah polusi udara, seperti konsentrasi PM<sub>10</sub> dan PM<sub>2.5</sub> adalah kerusakan pada manusia[7]. Untuk referensi web kami kemungkinan menggunakan streamlit , dikarenakan streamlit dapat digunakan untuk menampilkan data dari python[8]. Algoritma ELM dapat menghasilkan kinerja generalisasi yang baik dalam banyak kasus dan dapat belajar ribuan kali lebih cepat daripada algoritma pembelajaran populer konvensional[9]. (KELM) meningkatkan kekokohan mesin pembelajaran ekstrim (ELM) dengan mengubah data yang tidak dapat dipisahkan secara linier dalam ruang dimensi rendah menjadi data yang dapat dipisahkan secara linear[10]. Menurut metode pembelajaran ELM, konsep kernel fungsi diperkenalkan ke ELM saat pemetaan fitur fungsi  $h(x)$  tidak diketahui, sehingga ELM berbasis kernel (KELM) metode terbentuk. Oleh karena itu, KELM dapat dianggap sebagai perpanjangan nonlinier dari ELM[11]. Penanganan missing data dengan KNN diawali dengan menentukan sejumlah tetangga terdekat atau observasi terdekat yang disimbolkan dengan K, kemudian menghitung jarak terkecil dari setiap observasi yang tidak mengandung missing data[12]. Visualisasi peta ISPU[13]. Menurut (katadata) Regresi dipakai untuk mengamati pengaruh antara dua variabel atau lebih. Hubungan variabel ini diwujudkan dalam model

matematis. Artikel ini telah tayang di Katadata.co.id dengan judul "Regresi Adalah Metode Analisis, Ini Jenis dan Rumus Hitungannya"[14]. Metode regresi merupakan sebuah metode statistik yang melakukan prediksi menggunakan pengembangan hubungan matematis antara variabel, yaitu variabel dependen (Y) dengan variabel independen (X)[15]. Nilai akurasi minimum didefinisikan sebagai akurasi terendah yang diharapkan peta tematik yang diberikan dari hasil uji akurasi yang diamati dan digunakan berdasarkan risiko konsumen. Karena ukuran tingkat ketidakpastian dimasukkan ke dalam nilai akurasi minimum, ini adalah indeks yang berguna untuk membandingkan hasil akurasi tes menggunakan ukuran sampel yang berbeda digunakan dalam fungsi kerugian untuk membandingkan biaya maksimum relatif yang diharapkan dari metode klasifikasi alternatif[16].

### **1.3 Constraint**

#### **1.3.1 Aspek Keberlanjutan (*sustainability*)**

Sistem informasi kualitas udara untuk DKI Jakarta berdasarkan data Indeks Standar Pencemaran Udara (ISPU) bertujuan sebagai layanan informasi yang dapat digunakan terus menerus kedepannya dengan data terkini.

#### **1.3.2 Aspek Penggunaan (*usability*)**

Sistem informasi yang dibuat berbasis website yang didesain seperti layanan informasi agar mempermudah pengoperasian dalam mengakses informasi kualitas udara di DKI Jakarta berdasarkan data Indeks Standar Pencemaran Udara (ISPU).

### **1.4 Kebutuhan yang Harus Dipenuhi**

Berdasarkan masalah, latar belakang, dan analisis yang telah dipaparkan, maka kebutuhan yang harus dipenuhi dari solusi yang akan diajukan antara lain:

- a. Sistem dapat melakukan klasifikasi dan regresi data ISPU dengan algoritma *Machine Learning*.
- b. Sistem dapat memprediksi kualitas udara berdasarkan masing – masing partikel udara di DKI Jakarta berdasarkan data ISPU.
- c. Sistem dapat menghasilkan indeks kualitas udara berdasarkan proses klasifikasi.
- d. Sistem dapat meramalkan kualitas udara 30 hari kedepan.
- e. Sistem informasi hasil klasifikasi dan regresi data berbasis website.

- f. Sistem dirancang dapat dengan mudah ditingkatkan (*upgrade*) untuk memenuhi kebutuhan atau perkembangan di masa mendatang.

## 1.5 Tujuan

Berdasarkan kebutuhan yang harus dipenuhi, maka tujuan dari tugas akhir ini adalah membuat sistem informasi kualitas udara di DKI Jakarta yang diperoleh dari proses *machine learning* berdasarkan dataset ISPU. Kemudian membuat sistem yang dapat menampilkan informasi tersebut ke masyarakat berbasis *website*.

## 1.6 Solusi

### 1.6.1 Karakteristik Produk

#### a. Fitur Utama:

- 1) Mengolah dataset ISPU dengan menerapkan metode klasifikasi dan regresi menggunakan algoritma *machine learning*.
- 2) Menyediakan informasi kualitas udara DKI Jakarta berbasis *website*.

#### b. Fitur Dasar:

- 1) Melakukan proses regresi data ISPU menggunakan algoritma *machine learning*.
- 2) Melakukan proses klasifikasi data ISPU menggunakan algoritma *machine learning*.
- 3) Mendapatkan hasil peramalan data kualitas udara 30 hari kedepan dengan algoritma *machine learning*.
- 4) Menyajikan hasil keseluruhan proses *machine learning* ke dalam *website*.

#### c. Fitur Tambahan:

- 1) Tampilan maps yang menyediakan informasi kualitas udara pada 5 titik daerah DKI Jakarta.
- 2) Fitur untuk memilih visualisasi hasil klasifikasi atau regresi pada *website*.
- 3) Menyediakan informasi lengkap mengenai ISPU.
- 4) Menggunakan lebih dari satu metode *machine learning*.

#### d. Sifat Solusi yang Diharapkan:

- 1) *Machine learning* dapat dengan baik melakukan proses klasifikasi dan regresi.
- 2) *Machine learning* dapat menghasilkan peramalan kualitas udara 30 hari kedepan.

- 3) Masyarakat umum dapat dengan mudah mengakses informasi kualitas udara di DKI Jakarta.
- 4) Tidak membutuhkan perawatan yang terlalu intensif dan mengeluarkan biaya mahal.

### 1.6.2 Usulan Solusi

Berdasarkan konstrain dan karakteristik dari produk, maka terdapat 2 alternatif solusi yang dapat ditawarkan.

#### 1.6.2.1 Solusi 1

Sistem ini dirancang dengan desain yang mudah untuk digunakan oleh pengguna, sistem ini menyediakan informasi tentang klasifikasi dan prediksi berdasarkan proses regresi *machine learning*. Selain prediksi sistem juga dapat meramalkan kualitas udara 30 hari kedepan. Algoritma *machine learning* yang digunakan adalah *Extreme Learning Machine* (ELM) dan *Kernel Extreme Learning Machine* (K-ELM) Hasil dari klasifikasi, prediksi, dan peramalan akan divisualisasikan kedalam *website*.

Skenario penggunaan produk:

- a. Pengguna dapat mengakses *website* melalui browser pada perangkat yang terhubung dengan jaringan internet.
- b. Pengguna dapat melihat indeks kualitas udara beserta nilai partikel udara pada waktu tertentu pada maps yang tersedia.
- c. Pengguna dapat melihat kualitas udara 30 hari kedepan yang didapat dari hasil peramalan atau *forecasting*.
- d. Pengguna dapat memilih menu visualisasi klasifikasi atau regresi.
- e. Pengguna dapat memilih algoritma *machine learning* yang diinginkan untuk mendapatkan visualisasi regresi.
- f. Pengguna dapat memasukkan nilai kualitas udara secara manual dan mendapatkan hasil klasifikasinya.
- g. *Website* untuk sistem dibuat menggunakan *framework Streamlit*.

Stakeholder yang terlibat:

- a. Dosen pembimbing sebagai pencetus dari proyek tugas akhir ini.
- b. Kelompok Tugas Akhir Capstone sebagai pelaksana proyek.
- c. Dinas Lingkungan Hidup DKI Jakarta sebagai sumber dataset ISPU.

#### 1.6.2.2 Solusi 2

Sistem ini dirancang dengan desain yang mudah untuk digunakan oleh pengguna, sistem ini menyediakan informasi tentang klasifikasi dan prediksi berdasarkan proses regresi *machine learning*. Algoritma *machine learning* yang digunakan adalah *Extreme Learning Machine* (ELM). Hasil dari klasifikasi dan prediksi akan divisualisasikan kedalam *website* kedepan dan divisualisasikan pada aplikasi android

Skenario penggunaan produk:

- a. Pengguna dapat mengakses aplikasi melalui perangkat android yang terhubung dengan jaringan internet.
- b. Pengguna dapat melihat indeks kualitas udara berdasarkan klasifikasinya.
- c. Pengguna dapat melihat indeks kualitas udara dalam bentuk tabel pada aplikasi android.
- d. Aplikasi android dibuat menggunakan *React Native*.

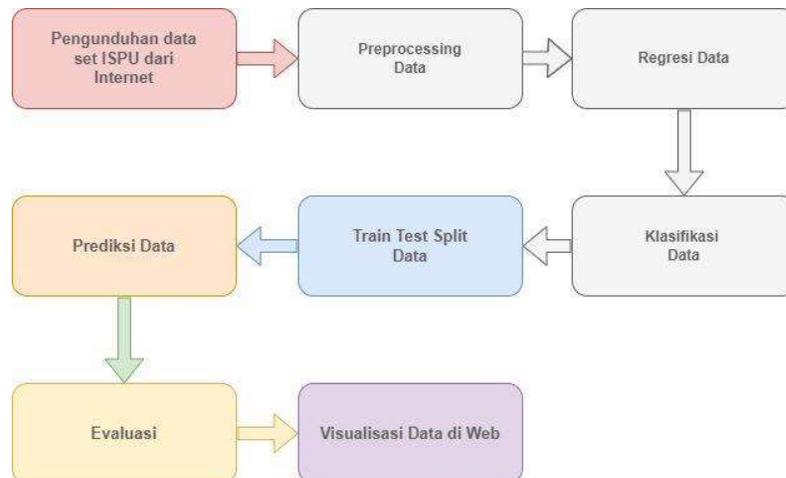
Stakeholder yang terlibat:

- a. Dosen pembimbing sebagai pencetus dari proyek tugas akhir ini.
- b. Kelompok Tugas Akhir Capstone sebagai pelaksana proyek.
- c. Dinas Lingkungan Hidup DKI Jakarta sebagai sumber dataset ISPU.

#### 1.6.3 Solusi yang Dipilih

Berdasarkan latar belakang dan *constrain* yang ada, solusi yang dipilih untuk project *Capstone Design* ini adalah solusi pertama. Penulis memilih solusi pertama yaitu membuat sistem yang menyediakan informasi tentang klasifikasi dan prediksi menggunakan metode *Extreme Learning Machine* (ELM) dan *Kernel Extreme Learning Machine* (K-ELM). Sistem juga meramalkan kualitas udara 30 hari kedepan menggunakan algoritma *machine learning*. hasil dari proses *machine learning* tersebut akan divisualisasikan kedalam *website*.

Sistem klasifikasi dan prediksi kualitas udara menggunakan metode ELM dan K-ELM berdasarkan Jakarta Open Data memiliki gambaran umum dimana terdapat langkah – langkah untuk mendapatkan hasil akhir. Langkah – langkah tersebut dapat dilihat pada diagram konseptual pada gambar 1.1 berikut ini.



**Gambar 1. 1 Diagram Konseptual**

Diagram konseptual diatas menjelaskan urutan proses untuk melakukan klasifikasi dan regresi menggunakan algoritma *machine learning* sampai hasil akhir yang ditampilkan kedalam *website*. Penjelasan setiap langkahnya seperti berikut:

1. Proses pertama dalam sistem ini adalah pengunduhan dataset ISPU pada website Jakarta Open Data dan Jakarta Rendah Emisi.
2. Setelah dataset diunduh, perlu dilakukan proses preprocessing pada dataset. Proses preprocessing pada dataset ISPU antara lain pengisian data kosong dan normalisasi data. Proses ini berguna untuk meningkatkan kinerja algoritma *machine learning*.
3. Dataset yang telah melalui proses *preprocessing* data dapat digunakan untuk proses klasifikasi dan regresi menggunakan algoritma *machine learning* dengan metode *Extreme Learning Machine* (ELM) dan *Kernel Extreme Learning Machine* (K-ELM).
4. Hasil dari proses klasifikasi dan regresi divisualisasikan kedalam *website*.

### 1.7 Kesimpulan dan Ringkasan CD-1

Proses klasifikasi dan prediksi kualitas udara di DKI Jakarta menggunakan metode *Extreme Learning Machine* (ELM) dan *Kernel Extreme Learning Machine* (K-ELM) berdasarkan dataset ISPU akan dilakukan pada *Capstone Project* ini. Tantangan yang dihadapi dalam proses ini adalah penggunaan metode *machine learning* yang terhitung baru dan jarang digunakan. Menurut literatur yang sudah dipelajari, metode *Extreme Learning Machine* (ELM) dan *Kernel Extreme Learning Machine* (K-ELM) memiliki keunggulan dalam hal klasifikasi dan regresi dibanding dengan metode lain yang sering digunakan. Oleh karena itu diperlukan banyak percobaan dalam *Capstone Project* ini.

Setelah proses *machine learning* selesai, hasil dari prediksi dan peramalan divisualisasikan ke dalam sebuah *website* dalam bentuk grafik. Di dalam *website*, pengguna dapat melihat 5 daerah DKI Jakarta yang memunculkan nilai untuk hasil prediksi, peramalan maupun klasifikasi. Untuk model klasifikasi juga dapat langsung diimplementasikan ke dalam sebuah *website*. Hal ini memungkinkan pengguna untuk dengan mudah mendapatkan hasil klasifikasi berdasarkan partikel udara yang dimasukkan secara manual. Dengan *website* yang dibuat, masyarakat dapat dengan mudah mendapatkan informasi kualitas udara di DKI Jakarta.