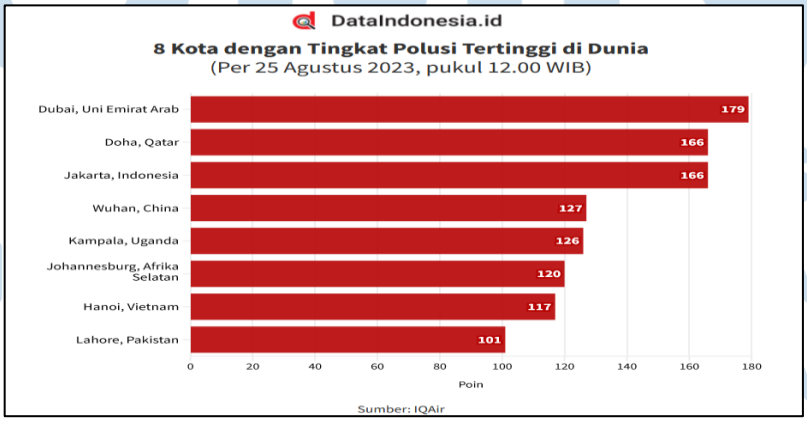
BAB I

PENDAHULUAN

1.1 Latar Belakang

Isu pencemaran udara dan kualitas udara menjadi bahan perbincangan yang hangat mendekati akhir tahun 2023. Pada rentang bulan Juli hingga Agustus 2023, lebih dari 34 ribu warganet media sosial *Twitter* mengeluhkan mengenai kondisi polusi udara dengan 85% perbincangan berasal dari Pulau Jawa [1]. Berdasarkan perbincangan tersebut, DKI Jakarta menjadi pusat perhatian terhadap topik pencemaran udara yang semakin memburuk. Pencemaran udara yang membuat kualitas udara semakin memburuk di DKI Jakarta bukanlah permasalahan baru, nyatanya dari tahun 2018 hingga 2022, rata-rata tingkat konsentrasi partikulat PM 2.5 pada provinsi tersebut mencapai 7 hingga 10 kali lipat lebih tinggi dari yang dianjurkan oleh WHO (*World Health Organization*) [2]. PM 2.5 merupakan salah satu dari sekian banyak polutan yang dapat mengganggu kesehatan dan meningkatkan risiko penyakit jantung hingga kanker paru-paru jika terpapar dalam jangka panjang [3]. Selain penyakit jantung dan kanker paru-paru, pencemaran udara juga dapat menyebabkan *asthma*, *bronchitis*, dan juga infeksi saluran pernapasan akut (ISPA) [4]. Misalnya, untuk kasus ISPA di DKI Jakarta per Agustus 2023 sendiri mengalami peningkatan mencapai 200 ribu kasus, jika dibandingkan pada masa pandemi *COVID-19* yang hanya mencapai sekitar 50 ribu [5].



Gambar 1. 1 Grafik Kota dengan Tingkat Polusi Tertinggi di Dunia

Gambar 1.1 [6] menunjukkan DKI Jakarta menempati urutan ke-3 sebagai kota dengan tingkat polusi tertinggi di dunia dimana skor indeks kualitas udara DKI Jakarta menyentuh angka 166 yang sama dengan Kota Doha di Qatar. Selanjutnya mengacu angka terdahulu pada tanggal 30 September 2019, DKI Jakarta menempati urutan ke-10 sebagai kota dengan kualitas udara terburuk di dunia dan indeks kualitas udara menyentuh angka 157 [7]. Angka tersebut memang tidak tetap dan dapat berubah-ubah, perbandingan antara tahun 2019 dan 2023 juga menunjukkan angka tersebut mengalami penurunan tetapi keduanya masih sama-sama berada di kategori tidak sehat. Dikutip dari laman *IQAir*, berdasarkan pengukuran indeks kualitas udara, angka 151 – 200 merupakan kategori tidak sehat dengan adanya kemungkinan meningkatnya dampak buruk terhadap kesehatan jantung ataupun paru-paru terutama untuk kelompok sensitif atau kelompok masyarakat yang memiliki kerentanan kesehatan terkait kualitas udara tersebut [8].

Indeks kualitas udara tersebut didapatkan melalui pengukuran konsentrasi polutan udara seperti konsentrasi partikulat udara seperti PM10 dan PM2.5 ataupun yang lainnya. Semakin tinggi konsentrasi menandakan semakin buruk kualitas udara. Selanjutnya, pengukuran kualitas udara di Indonesia terdapat pada Indeks Standar Pencemar Udara (ISPU) yang ditetapkan oleh Kementerian Lingkungan Hidup dan Kehutanan (KLHK). Kedua indeks tersebut sama-sama mengukur konsentrasi polutan udara yang dihasilkan oleh banyak hal salah satunya adalah aktivitas manusia seperti penggunaan kendaraan bermotor dan kegiatan industri. Pengukuran yang dilakukan oleh pemerintah DKI Jakarta sendiri mengikuti ketetapan dari Keputusan Menteri Negara Lingkungan Hidup No. 45 Tahun 1997 sebagaimana terdapat pada pasal 3 mengenai kegunaan Indeks Standar Pencemar Udara (ISPU) sebagai bahan informasi kepada masyarakat tentang kualitas udara ambien di lokasi dan waktu tertentu, serta pada pasal 4 mengenai daftar parameter untuk ISPU tersebut yang terdiri dari partikulat (PM10), karbon monoksida (CO), sulfur dioksida (SO2), nitrogen dioksida (NO2), dan ozon (O3) [9]. Kemudian pemerintah mengeluarkan Peraturan Menteri Lingkungan Hidup dan Kehutanan nomor 14 tahun 2020 dan menambah parameter partikulat (PM2.5) dan juga hidrokarbon (HC) yang dipantau dengan bantuan Stasiun Pemantau Kualitas Udara

Ambien (SPKUA) [10]. Selanjutnya, data-data observasi pengukuran polutan dapat dilihat oleh masyarakat melalui aplikasi atau *platform* terkait pengukuran kualitas udara.

Adanya informasi tersebut dapat membantu masyarakat mengetahui jenis-jenis polutan udara sebagai sumber tingginya angka polusi udara di DKI Jakarta berdasarkan nilai ISPU hariannya. Oleh karena itu diperlukan penerapan *data mining* untuk dapat menemukan pola terhadap polusi udara yang ada di DKI Jakarta menggunakan data Indeks Standar Pencemar Udara (ISPU) terutama pada polutan yang sering menjadi pencemar utama. Untuk dapat menemukan pola yang tersembunyi pada data ISPU DKI Jakarta, serta mengelompokkan data berdasarkan Stasiun Pemantau Kualitas Udara (SPKU) yang ditempatkan di wilayah berbeda, dibutuhkan metode *clustering* karena dapat melakukan identifikasi pola temporal pada polutan udara tersebut [11]. Selain itu, alasan lainnya dilakukan *clustering* adalah karena jumlah data yang banyak tidak cocok untuk dilakukan analisis secara tradisional, serta informasi yang didapatkan akan lebih bermakna [12]. Kemudian metode *clustering* tersebut digabungkan menggunakan pendekatan *hybrid* dengan model *neural network Long Short-Term Memory* (LSTM) sebagai model prediksi konsentrasi polutan udara dari data ISPU. Tujuannya adalah menemukan pola dengan cara memprediksi pada masa yang akan datang, misalnya 14 hari ke depan. Adanya metode *data mining* dengan menggunakan gabungan *clustering* dan juga prediksi dapat mempermudah pencarian pola pada data ISPU tersebut.

Metode *clustering* sendiri merupakan suatu metode untuk memproses suatu data dan membaginya menjadi ke dalam kelompok-kelompok atau dinamakan *cluster* berdasarkan pola tertentu [13]. Selanjutnya, penerapan *LSTM* sendiri merupakan bagian dari *neural network* yang dapat memproses data lebih kompleks dan saling terhubung dibandingkan model *machine learning* seperti *clustering* karena menggunakan konsep otak manusia dalam mengolah informasi [14]. Penelitian ini akan menggunakan *hybrid clustering* dimana menggunakan gabungan *K-Means* dan *K-Medoids* berdasarkan penelitian di Iraq [15]. Untuk penelitian mengenai LSTM sendiri ada pada penelitian untuk memprediksi konsentrasi PM10 berdasarkan data Stasiun Pemantau Kualitas Udara di Malaysia menggunakan

hybrid model antara *K-Means* dan LSTM dengan *output* model yang dihasilkan tersebut lebih efektif dalam penerapan di kehidupan nyata [16]. Pada penelitian ini sendiri akan menggunakan *hybrid model* gabungan dari *K-Means* dan *K-Medoids* sehingga nantinya akan muncul *cluster* akhir dan setelah itu diprediksi menggunakan *Long Short-Term Memory* (LSTM) dan juga menggunakan arahan dari penelitian [16] terkait data yang akan digunakan seperti data cuaca. Oleh karena itu, data yang akan digunakan pada penelitian ini adalah Indeks Standar Pencemar Udara (ISPU) yang berasal dari Dinas Lingkungan Hidup DKI Jakarta dari tahun 2010 sampai dengan 2023 dengan adanya data pendukung berupa data cuaca

Untuk tujuan dari penelitian ini adalah melakukan implementasi *hybrid clustering* menggunakan *K-Means* dan *K-Medoids* yang kemudian menggunakan *LSTM* dalam memprediksi data-data ISPU dan data pendukugnya. Penelitian ini juga memiliki tujuan untuk melihat performa hasil prediksi *LSTM* berdasarkan hasil *hybrid clustering* sebagai *input* dengan menggunakan nilai *RMSE* untuk setiap *model LSTM* yang telah dibuat di masing-masing *optimizer*, yaitu *Adam* dan *SGD*. Kemudian, dicari hasil terbaik dengan melihat nilai evaluasi *RMSE* terkecil antara *optimizer* pada setiap *model* yang kemudian dilakukan visualisasi akhir berupa prediksi. Hasil akhir tersebut adalah melihat pola antara data prediksi dengan data aktual dan juga terdapat hasil *forecast* selama 14 hari ke depan. Kemudian, hasil penelitian juga dapat membantu untuk menambah informasi mengenai penerapan *data mining* dengan menggunakan metode *clustering* dan prediksi pada *dataset* ISPU. Selain itu, diharapkan penelitian ini juga memberikan wawasan mengenai Indeks Standar Pencemar Udara (ISPU) sebagai acuan pengukuran kualitas udara agar lebih memahami kondisi udara di lingkungan sekitar.

1.2 Rumusan Masalah

Dari latar belakang di atas dapat diambil beberapa rumusan masalah seperti berikut:

1. Bagaimana penerapan *data mining* berupa *hybrid model* menggunakan gabungan algoritma *K-Means* dan *K-Medoids* dengan *Long Short-Term*

Memory (LSTM) dalam memprediksi data Indeks Standar Pencemar Udara (ISPU) DKI Jakarta, serta data cuaca sebagai pendukung?

2. Bagaimana *output* atau performa yang dihasilkan oleh *hybrid model* gabungan *K-Means* dan *K-Medoids* dengan *Long Short-Term Memory* (LSTM) pada 2 *optimizer* yang berbeda?
3. Bagaimana visualisasi data prediksi dari hasil akhir *hybrid model* terbaik berdasarkan nilai evaluasi *RMSE*, serta hasil *forecast* selama 14 hari ke depan?

1.3 Batasan Masalah

Batasan masalah pada penelitian ini bertujuan agar pembahasan dan alur penelitian tidak mengarah ke luar *scope* permasalahannya. Oleh karena itu, batasan masalah pada penelitian ini adalah sebagai berikut:

1. Penelitian ini terbatas pada *clustering* dan prediksi menggunakan *hybrid model* terhadap data Indeks Standar Pencemar Udara (ISPU) DKI Jakarta.
2. Data yang digunakan pada penelitian ini diperoleh dari Dinas Lingkungan Hidup DKI Jakarta dari 1 Januari 2010 sampai dengan 30 November 2023 melalui *website* Satu Data Jakarta (https://satudata.jakarta.go.id/), serta data pendukung terkait data cuaca seperti suhu, kelembapan, dan kecepatan angin yang diperoleh melalui *website* (https://www.visualcrossing.com/).
3. Pembagian *dataset* terbagi menjadi ISPU 2010 – 2023 dan juga ISPU 2021 – 2023 karena terdapat parameter pengukuran baru, yaitu PM2.5 untuk tahun 2021 dan seterusnya.
4. Untuk penerapan *hybrid model*, model *clustering* yang digunakan adalah gabungan dari *K-Means* dan *K-Medoids*, sedangkan untuk model prediksi adalah menggunakan *Long Short-Term Memory* (LSTM).
5. Variasi *dataset* sebagai *input* untuk model *clustering* adalah menggunakan ISPU 2010 – 2023 yang dibagi menjadi secara keseluruhan dan *training-test set*, serta *dataset* ISPU 2021 – 2023. Variasi untuk *model* prediksi sendiri adalah berdasarkan hasil *clustering*, keseluruhan *dataset*, dan per stasiun untuk masing-masing ISPU 2010 – 2023 dan ISPU 2021 – 2023. Pembagian *input* pada prediksi tetap menggunakan *training-test set* dengan

rasio 80:20. Fokus dari *model* prediksi adalah menggunakan hasil *clustering* dan menggunakan variabel ISPU, selain dari variabel ISPU hanya bersifat opsional sebagai tambahan informasi.

6. Hasil visualisasi yang ditampilkan hanya menggunakan *model* prediksi per *cluster* dengan nilai *RMSE* terkecil. Selain itu, untuk *model LSTM* secara *multivariate* akan ditampilkan secara keseluruhan sesuai dengan nilai *RMSE* terkecil. Masing-masing visualisasi juga terdapat hasil *forecast* selama 14 hari ke depan.

1.4 Tujuan dan Manfaat Penelitian

1.4.1 Tujuan Penelitian

Berdasarkan latar belakang permasalahan, rumusan masalah dan batasan masalah, ditentukan tujuan pada penelitian ini yang terdiri dari:

1. Untuk mengetahui implementasi dari *Hybrid Model* menggunakan *K-Means* dan *K-Medoids* dengan *model* prediksi *Long Short-Term Memory* (*LSTM*) terhadap *dataset* Indeks Standar Pencemar Udara (ISPU) DKI Jakarta serta data pendukung berupa data cuaca.
2. Untuk mengetahui nilai evaluasi yang terbaik dari 2 *optimizer* yang digunakan pada *model* prediksi *Long Short-Term Memory* (LSTM)
3. Untuk mengetahui hasil visualiasi berupa perbandingan data asli dengan data prediksi berdasarkan evaluasi terbaik dari penerapan *hybrid model* antara *clustering* dan prediksi, serta hasil *forecast* selama 14 hari ke depan.

1.4.2 Manfaat Penelitian

Manfaat dari penelitian ini adalah sebagai berikut:

1. Memberikan pemahaman baru dari penelitian yang sudah ada sebelumnya terutama pada *data mining* dengan metode *clustering* atau prediksi dari adanya *hybrid model* yang sudah ditentukan.
2. Memberikan informasi mengenai pencemar utama yang sering menyumbang angka ISPU tertinggi dengan adanya analisis terhadap data ISPU di DKI Jakarta dari tahun 2010 hingga 2023.

3. Memberikan informasi berupa visualisasi hasil prediksi data ISPU beserta *forecast* selama 14 hari ke depan.

1.5 Sistematika Penulisan

Laporan ini memiliki sistematika penulisan agar lebih jelas dan terstruktur. Berikut ini adalah sistematika penulisan laporan tersebut:

BAB I PENDAHULUAN

Bagian laporan ini menjelaskan latar belakang penelitian, rumusan masalah, batasan masalah, tujuan dan manfaat penelitian, serta sistematika penulisan laporan penelitian.

BAB II LANDASAN TEORI

Bagian laporan ini menjelaskan teori pendukung dalam melakukan penelitian yang dapat berupa artikel jurnal dan buku.

BAB III METODOLOGI PENELITIAN

Bagian laporan ini menjelaskan gambaran umum objek penelitian, metode yang digunakan dalam melakukan penelitian, hingga pengumpulan data penelitian.

BAB IV ANALISIS DAN HASIL PENELITIAN

Bagian laporan ini menjelaskan analisis masalah yang dilakukan dan juga hasil dari penelitian yang sudah dilakukan tersebut.

BAB V SIMPULAN DAN SARAN

Bagian laporan ini menjelaskan kesimpulan dari keseluruhan penelitian yang sudah dilakukan.

UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA