# **BAB II**

# **TINJAUAN LITERATUR**

# **Landasan Teori**

# **Data Spatiotemporal**

Data spatiotemporal terdiri dari representasi spasial dan temporal yang mengacu pada data yang dikumpulkan dan dianalisis melintasi kedua dimensi ruang dan waktu (Hamdi et al., 2022). Data spasial merepresentasikan koordinat, titik, garis, poligon sederhana, dan objek 3D. Data temporal berkaitan dengan waktu baik masa lalu, masa kini, maupun masa depan. Data spatiotemporal mewakili fenomena yang berbeda mulai dari skala mikro DNA dan evolusi sel, hingga yang global, misalnya, perubahan iklim, curah hujan, sebaran penyakit, dan sebagainya (Yang et al., 2020). Analisis data spatiotemporal memiliki kompleksitas tinggi dan digunakan dalam berbagai bidang seperti meteorologi, geografi, ilmu lingkungan, kedokteran, transportasi, dan sebagainya (Amran et al., 2020). Metode *data mining* dapat digunakan untuk mengolah dan menganalisis data spatiotemporal untuk berbagai tujuan, seperti prediksi kualitas udara, pengukuran nilai tanah, dan ramalan cuaca (Asgari et al., 2022)(Alizanovic, 2023). Selain itu, visualisasi data spatiotemporal juga digunakan untuk melakukan pemetaan sebaran penyakit dan analisis geografis lainnya (Heldayani et al., 2021)(Lasari et al., 2023).

# **Indeks Standar Pencemar Udara (ISPU)**

Indeks kualitas udara merupakan suatu parameter tanpa satuan yang memberikan informasi mengenai gambaran kondisi kualitas udara ambien di lokasi dan waktu tertentu kepada masyarakat. Indeks kualitas udara yang resmi digunakan di Indonesia berdasarkan Keputusan Menteri Negara Lingkungan Hidup Nomor 45 Tahun 1997 adalah Indeks Standar Pencemar Udara (ISPU) (Eko H., 2017). Penilaian ISPU didasarkan pada dampak yang dapat ditimbulkan terhadap kesehatan manusia, nilai estetika lingkungan, dan makhluk hidup lainnya. ISPU dapat dijadikan sebagai sistem peringatan dini (*early warning system*) bagi masyarakat yang tinggal di daerah rawan terdampak kebakaran hutan dan lahan. ISPU juga digunakan sebagai bahan pertimbangan untuk merumuskan upaya-upaya pengendalian pencemaran udara, baik oleh pemerintah pusat maupun daerah.

Terdapat tujuh parameter pencemar udara yang digunakan dalam perhitungan ISPU dan didasari pada besarnya risiko terhadap kesehatan manusia, yaitu PM10, PM2.5, NO2, SO2, CO, O3, dan HC (Chaniago et al., 2020). ISPU dihitung berdasarkan pada nilai-nilai ISPU batas atas dan bawah, konsentrasi ambien batas atas dan bawah, serta konsentrasi ambien dari hasil pengukuran. Persamaan perhitungan ISPU sebagai berikut:

I = ISPU terhitung

Ia = ISPU batas atas

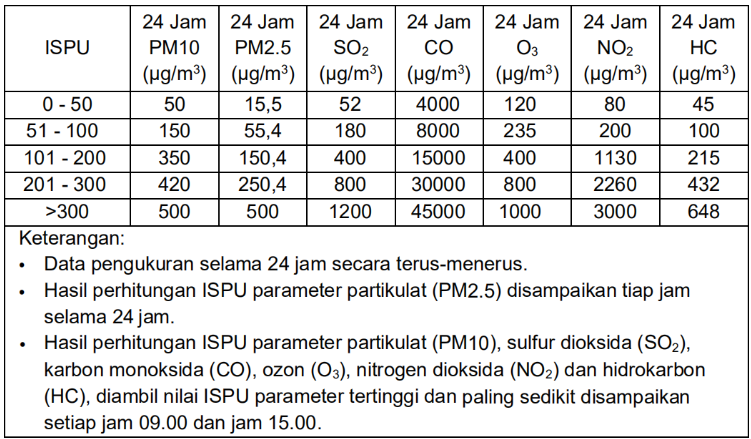
Ib = ISPU batas bawah

Xa = Konsentrasi ambien batas atas

Xa = Konsentrasi ambien batas bawah

Xx = Konsentrasi ambien nyata hasil pengukuran

Nilai batas atas ISPU paling tinggi adalah lebih dari 300, sedangkan nilai batas bawah ISPU paling rendah adalah 50. Kosentrasi ambien batas atas dan bawah berbeda pada tiap parameter dan diperoleh dari tabel konversi nilai konsentrasi parameter ISPU. Nilai konsentrasi ambien yang sebenarnya diperoleh dari rata-rata konsentrasi ambien selama 24 jam pengukuran (Firman et al., 2023). Tabel konversi nilai konsentrasi parameter ISPU sebagai berikut:



Berdasarkan Kementerian Lingkungan Hidup dan Kehutanan, ISPU terbagi menjadi lima kategori sesuai rentang nilainya (Wibawana, 2023), yaitu:

1. Kategori Baik (0 – 50)

ISPU dengan rentang nilai 0 – 50 memiliki kategori Baik. Tingkat kualitas udara masih sangat baik dan tidak memberikan efek negatif terhadap kesehatan manusia dan hewan, serta tidak berpengaruh terhadap tumbuhan, bangunan, dan nilai estetika.

1. Kategori Sedang (51 – 100)

ISPU dengan rentang nilai 51 – 100 memiliki kategori Sedang. Tingkat kualitas udara masih dapat diterima oleh kesehatan manusia dan hewan, tetapi sedikit berpengaruh terhadap tumbuhan yang sensitif dan nilai estetika.

1. Kategori Tidak Sehat (101 – 200)

ISPU dengan rentang nilai 101 – 200 memiliki kategori Tidak Sehat. Tingkat kualitas udara bersifat merugikan pada manusia maupun hewan atau dapat menimbulkan kerusakan pada tumbuhan dan nilai estetika.

1. Kategori Sangat Tidak Sehat (201 – 300)

ISPU dengan rentang nilai 201 – 300 memiliki kategori Sangat Tidak Sehat. Tingkat kualitas udara dapat menyebabkan kerugian dan meningkatkan risiko kesehatan pada beberapa kelompok populasi yang terpapar.

1. Kategori Berbahaya (300+)

ISPU dengan rentang nilai lebih dari 300 memiliki kategori Berbahaya. Tingkat kualitas udara berbahaya dan secara umum dapat menyebabkan kerugian kesehatan serius pada populasi dan memerlukan penanganan dengan cepat.

Data ISPU diperoleh dari pengoperasian Stasiun Pemantauan Kualitas Udara Ambien dan disampaikan kepada masyarakat setiap 24 jam dari data sebelumnya dan berlaku 24 jam ke depan. Waktu pengambilan data terakhir dilakukan pada pukul 15.00 WIB. ISPU dapat diakses melalui aplikasi ISPU Net yang memungkinkan masyarakat mengetahui kondisi kualitas udara secara *real-time* di seluruh wilayah (Ramadhan P, 2021).

# ***Machine Learning***

*Machine Learning* merupakan sebuah teknologi yang memungkinkan mesin untuk belajar dan membuat keputusan secara otomatis tanpa instruksi pengguna. *Machine Learning* melalui tahap pembelajaran menggunakan data pelatihan khusus untuk mengotomatiskan proses pembuatan model analitis, yang memungkinkan model untuk membuat keputusan yang tepat dan menyelesaikan tugas-tugas terkait (Janiesch et al., 2021). Terutama pada tugas yang berkaitan dengan data *high-dimensional*, seperti klasifikasi, *clustering*, dan regresi, *Machine Learning* menunjukkan penerapan model yang baik. Teknologi ini digunakan dalam berbagai bidang, seperti pendidikan, ekonomi, teknologi, sosial, dan lainnya. Sudah banyak penerapan algoritma *machine learning* yang telah dilakukan seperti deteksi penipuan (Gupta et al., 2022), pengenalan ucapan dan gambar (Jimenez-Mesa et al., 2023), pemrosesan bahasa alami (NLP), dan lainnya (Janiesch et al., 2021).

Terdapat tiga tipe utama *Machine Learning* yang umum digunakan, yaitu *Supervised Learning*, *Unsupervised Learning*, dan *Reinforcement Learning*. *Supervised Learning* adalah tipe *Machine Learning* di mana algoritma dilatih menggunakan kumpulan data yang diberi label, dengan setiap titik data memiliki label yang sesuai (Mobarak et al., 2023). Hal ini menyiratkan bahwa beberapa data telah ditandai dengan jawaban yang benar. Tipe ini diterapkan untuk model klasifikasi dan regresi. Beberapa algoritma yang umum digunakan dalam *Supervised Learning*, yaitu *Support Vector Machine* (SVM), Naïve Bayes, K-NN, *Random Forest*, *Linear Regression*, De*cision Tree*, dan lainnya (Allenbrand, 2023).

Berbeda dengan *Supervised Learning*, *Unsupervised Learning* tidak memerlukan data pelatihan yang diberi label. Tipe ini menganalisis serta mempelajari struktur dan pola secara ekslusif dari data yang tidak berlabel dan sudah ada sebelumnya, kemudian mengungkap informasi tersembunyi dari data yang diberikan (Mobarak et al., 2023). Tipe ini diterapkan untuk model *clustering*, deteksi anomali, dan reduksi dimensi (Zipfel et al., 2023). Berapa algoritma yang digunakan dalam *Unsupervised Learning*, yaitu K-Means, Fuzzy, C-Means, dan lainnya.

*Reinforcement Learning* adalah tipe *Machine Learning* di mana agen belajar berinteraksi dengan lingkungannya melalui *trial and* *error*, lalu menerima umpan balik dalam bentuk *reward* atau *punishment*. Umpan balik digunakan untuk menyesuaikan perilaku agen dan meningkatkan kinerjanya seiring waktu, Tujuan dari pembelajaran ini adalah mempelajari bagaimana mengambil tindakan untuk memaksimalkan *reward* yang didapatkan (Pichka, 2023). *Reinforcement Learning* telah diterapkan dalam berbagai bidang, seperti robotik, NLP, *game*, dan lainnya (Mandlekar et al., 2019) (Wang, 2021).

# **Prediksi**

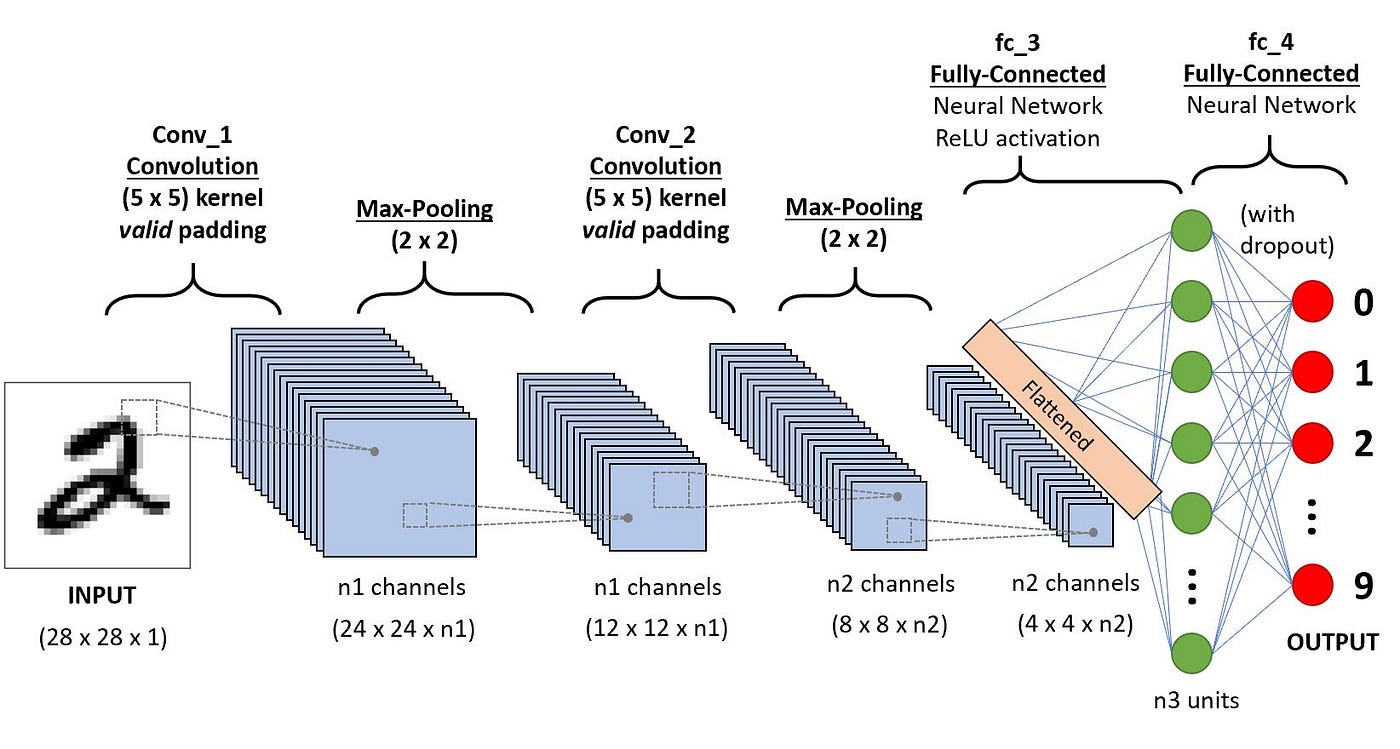
Prediksi adalah salah satu metode dalam *Supervised Learning* untuk memperkirakan *output* di masa yang akan datang melalui proses pembelajaran atau pelatihan. Pembelajaran ini dilakukan berdasarkan data *input* dari masa lalu dan kiniyang telah diberikan dengan tujuan untuk meminimalkan kesalahan yang terjadi (Dewi et al., 2022). Data prediksi dapat berasal dari berbagai sumber, seperti data historis atau data yang terus-menerus diperbarui (Hartatik et al., 2023). Data ini kemudian digunakan untuk melatih model agar dapat mempelajari pola antara *input* dan *output*, sehingga dapat melakukan prediksi terhadap data baru.

# ***Oversampling***

*Oversampling* merupakan sebuah teknik dalam *machine learning* untuk mengatasi permasalahan *imbalanced data.* Kondisi *imbalanced* *data* dapat memengaruhi kinerja model di mana hasil evaluasi menjadi bias. Hal ini dapat terjadi karena data pada kelas mayoritas yang sangat banyak, sedangkan pada kelas minoritas sangat sedikit, atau kedua kelas muncul secara bersamaan (Triyanto & Kusumaningrum, 2017). Teknik *oversampling* menyeimbangkan dataset dengan meningkatkan ukuran sampel dari kelas minoritas. Peningkatan ukuran sampel dilakukan dengan cara membuat data buatan sebanyak yang diperlukan dari kelas minoritas, sehingga jumlah sampel dari kelas minoritas menjadi seimbang dengan kelas mayoritas (Qadrini et al., 2022). Salah satu algoritma *oversampling* yang umum digunakan adalah *Synthetic Minority Over-sampling Technique* (SMOTE). Algoritma ini mampu menangani permasalahan *imbalanced class* pada dataset karena mengurangi *overfitting* dan menghasilkan akurasi yang baik.

# ***Convolutional Neural Network* (CNN)**

*Convolutional Neural Network* (CNN) merupakan salah satu algoritma *deep learning* yang termasuk dalam jaringan *feed forward* di mana informasi mengalir searah dari *input* ke *output*. CNN dirancang untuk tugas pengolahan citra baik gambar, ucapan, atau sinyal audio, seperti klasifikasi, deteksi, segmentasi, dan prediksi (Keita, 2023). Pada beberapa penelitian, CNN juga digunakan pada data *time series* (Kim et al., 2023) dan spatiotemporal (Zhang et al., 2020). CNN terdiri dari lapisan-lapisan utama, yaitu *convolutional layer*, *ReLU layer*, *pooling layer*, dan *fully connected layer.* Arsitektur CNN (Keita, 2023) dapat dilihat pada gambar x.



Input pada *convolutional layer* adalah sebuah gambar dengan ukuran di mana adalah tinggi dan lebar pada gambar dan adalah jumlah saluran atau kedalaman (sesuai dengan RGB dalam gambar, jika hitam putih = 1 dan berwarna = 3). Pada *convolutional layer*, juga terdapat pendeteksi fitur yang dikenal sebagai kernel atau filter dengan ukuran di mana lebih kecil dibandingkan dimensi gambar dan dapat sama dengan jumlah saluran atau lebih kecil dan mungkin berbeda pada tiap kernel. Kernel akan memeriksa apakah fitur tersebut ada pada gambar dengan melintasi bidang reseptif gambar, yang disebut dengan konvolusi. Operasi konvolusi dilakukan berulang kali hingga kernel mencakup seluruh bagian gambar dengan ukuran filter berupa matriks 3x3. Hasil proses konvolusi disebut *feature map* yang dapat dilihat pada gambar x.



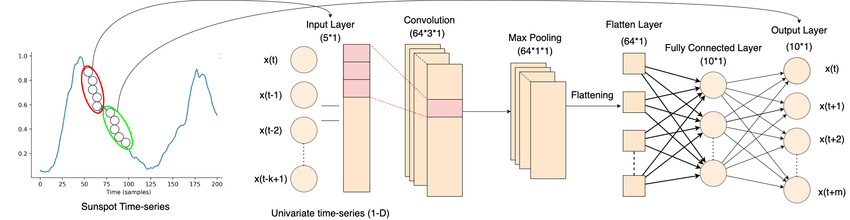
*Pooling layer* bertujuan untuk mengurangi jumlah parameter (*spatial resolution*) pada *feature map* yang disebut operasi *spatial pooling*. Pada *layer* ini, kernel mengisi *output* *array* dengan menerapkan fungsi agregasi dalam bidang reseptif. Terdapat dua jenis *spatial pooling*, yaitu *average pooling* dan *max pooling*. Perbedaan kedua jenis *pooling* ini ditunjukan pada gambar x. Matriks dari *feature map* dibagi-bagi dengan ukuran . Pada *average pooling*, filter menghitung nilai rata-rata untuk dikirim ke *output array*. Sedangkan pada *max pooling*, filter memilih nilai maksimum dari tiap bagian untuk dikirim ke *output array*. Ekstraksi fitur ini mengurangi risiko *overfitting* dan kompleksitas, serta meningkatkan efisiensi model.



Pada *fully connected layer*, setiap node pada *output layer* terhubung langsung ke node sebelumnya dengan *input* matriks satu dimensi. CNN menerapkan fungsi aktivasi ReLU (*Rectified Linear Unit)* untuk non-linieritas pada model. Lapisan ini melakukan tugas seperti klasifikasi dan prediksi berdasarkan fitur yang telah diekstrak sebelumnya dengan filter yang berbeda menggunakan *softmax*.

# **CNN untuk Time-series**

1. Input Layer yang berukuran 𝑁×𝑘, dimana 𝑘 merupakan jumlah data time-series univariate yang menjadi masukan dan 𝑁 merupakan panjang dari data time-series tersebut.
2. Convolutional Layer dengan jumlah filter sebanyak 𝑚 dan ukuran filter-filter tersebut 𝑙×𝑘, dimana 𝑘 merupakan jumlah data time-series univariate dari input layer dan 𝑙 merupakan panjang dari filter tersebut. Nilai dari 𝑚 dan 𝑙 merupakan parameter yang dapat diatur.
3. Pooling Layer yang sama dengan pooling layer pada arsitektur CNN pada umumnya, antara max pooling atau average pooling
4. Feature Layer dimana hasil semua operasi convolution dan pooling terhadap data time-series multivariate disambung menjadi sebuah time-series univariate baru yang akan menjadi input untuk fully connected layer.
5. Output Layer merupakan layer yang memiliki 𝑛 buah neuron, dimana 𝑛 jumlah kelas yang ada untuk timeseries classification terkait.



# **ST-ResNet**

# **Mean Squared Error (MSE) dan Mean Absolute Error (MAE)**

# **Studi Literatur**