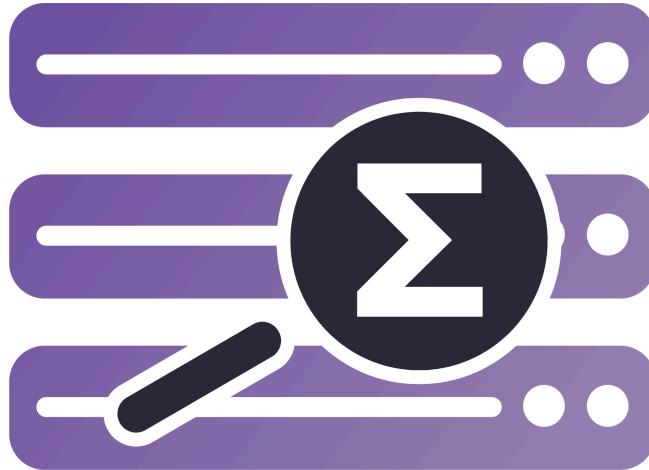




LAPORAN PENULISAN

Peramalan Pemadaman Listrik Akibat Cuaca Ekstrem Berdasarkan Metode *Weighted Ensemble*



ID Tim : SSDS2024030
Nama Tim : Martabak Sukarasa
Anggota :

1	Marcell Christian Franky	215090401111006
2	German Mindo Simarmata	215090400111015
3	Muhammad Faizal	215090400111039

SEBELAS MARET STATISTIC DATA ANALYTICS

SEBELAS MARET STATISTIC FAIR

2024





KATA PENGANTAR

Puji dan syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa atas rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan laporan dengan judul "Peramalan Pemadaman Listrik Akibat Cuaca Ekstrem Berdasarkan Metode *Weighted Ensemble*". Laporan ini disusun sebagai bagian dari final lomba Sebelas Maret Statistics Data Science (SSDS) yang merupakan kompetisi data sains tingkat nasional sebagai bagian dari acara Sebelas Maret Statistics Fair (SSF). Laporan ini disusun dalam rangka mengembangkan solusi prediksi yang akurat terhadap dampak cuaca ekstrem pada jaringan listrik.

Pemadaman listrik akibat fenomena alam telah menjadi tantangan di berbagai wilayah, terutama ketika menghadapi kondisi cuaca ekstrem. Kondisi ini berdampak luas pada kehidupan masyarakat serta sektor ekonomi. Dengan demikian, pentingnya upaya mitigasi dan strategi pencegahan menjadi perhatian utama dalam berbagai bidang, termasuk pengelolaan infrastruktur kelistrikan. Laporan ini berfokus pada pengembangan solusi prediksi yang diharapkan dapat membantu meningkatkan kesiapsiagaan dan ketahanan jaringan listrik terhadap cuaca ekstrem di masa mendatang.

Penulis menyadari bahwa laporan ini tidak akan terselesaikan tanpa bantuan, bimbingan, serta dukungan dari berbagai pihak. Oleh karena itu, penulis ingin mengucapkan terima kasih kepada seluruh pihak yang telah membantu dalam penyusunan laporan ini. Akhir kata, penulis berharap laporan ini dapat memberikan kontribusi yang berarti bagi perkembangan penelitian dan solusi penerapan teknologi dalam penanganan masalah cuaca ekstrem terhadap infrastruktur kelistrikan. Penulis juga berharap agar laporan ini dapat bermanfaat bagi pembaca dan pihak yang berkepentingan.

Malang, 22 Oktober 2024

Penulis





DAFTAR ISI

KATA PENGANTAR.....	2
DAFTAR ISI.....	3
BAB 1 PENDAHULUAN.....	4
A. Latar Belakang.....	4
B. Rumusan Masalah.....	5
C. Tujuan.....	5
BAB 2 METODOLOGI.....	6
A. Deskripsi Dataset.....	6
B. Metode Analisis.....	8
1. Analisis Data Eksploratori.....	8
2. Data Preprocessing.....	8
BAB 3 HASIL DAN PEMBAHASAN.....	11
A. Analisis.....	11
B. Pembahasan.....	13
BAB 4 KESIMPULAN.....	19
A. Kesimpulan.....	19
DAFTAR PUSTAKA.....	20





BAB 1 PENDAHULUAN

A. Latar Belakang

Perubahan iklim adalah perubahan pola iklim yang disebabkan oleh emisi gas rumah kaca. Emisi gas rumah kaca membuat panas terperangkap di atmosfer bumi dan menjadi pendorong utama pemanasan global (Fawzy et al, 2020). Pemanasan global memicu kejadian cuaca ekstrem yang makin intens, seperti gelombang panas, kekeringan, dan curah hujan tinggi yang terjadi setiap tahun. Frekuensi cuaca ekstrem ini memengaruhi manusia, infrastruktur, dan alam (Clarke et al, 2022). Salah satu infrastruktur yang terdampak adalah infrastruktur kelistrikan (Anel, 2024).

Beberapa negara merasakan dampak lumpuhnya infrastruktur kelistrikan karena cuaca ekstrem. Pada tahun 2021, Amerika Utara dan Spanyol mengalami badai musim dingin parah yang menyebabkan runtuhnya jaringan listrik. Selain itu, pada tahun 2024, Vietnam dilanda topan Yagi yang menghentikan operasional jaringan listrik di beberapa daerah di Vietnam dan China. Lumpuhnya jaringan listrik di negara tersebut berdampak pada kerugian sektor ekonomi hingga miliaran dolar Amerika Serikat (Anel, 2024). Peristiwa ini menekankan pentingnya strategi mitigasi pemadaman listrik untuk meminimalkan dampak cuaca ekstrem.

Sistem peringatan dini yang akurat sangat penting untuk memitigasi risiko pemadaman listrik melalui langkah-langkah preventif seperti perbaikan infrastruktur dan alokasi sumber daya. Strategi mitigasi untuk mengurangi dampak cuaca ekstrem pada infrastruktur kelistrikan adalah dengan metode prediksi yang didasarkan pada indikator cuaca (Panteli dan Mancarella, 2015). Indikator cuaca seperti kecepatan angin, tekanan atmosfer, kelembapan, dan curah hujan yang memiliki peran penting dalam memprediksi pemadaman listrik akibat cuaca ekstrem. Pemanfaatan indikator cuaca dapat menggunakan berbagai metode prediksi (Bollinger dan Dijkema, 2016).

Metode statistik adalah salah satu metode prediksi konvensional berbasis rata-rata klimatologis (Panteli dan Mancarella, 2015). Namun, metode statistik cenderung kurang responsif terhadap perubahan pola cuaca yang dinamis dan tidak mampu menangkap kompleksitas hubungan antara variabel cuaca dan risiko pemadaman. Metode alternatif lain adalah dengan metode *Machine Learning* (ML) (Anochi, 2021). ML menawarkan pendekatan yang lebih efektif karena kemampuannya untuk menganalisis data besar dengan cepat, mengidentifikasi pola kompleks, dan beradaptasi dengan perubahan pola cuaca (Priasni dan Oswari, 2021). Beberapa algoritma yang sering digunakan dalam kasus ini termasuk *Decision Tree* (DT), *Random Forest* (RF), *Gradient Boosting* (GB), dan *Support Vector Machine* (SVM). Namun, setiap algoritma memiliki keunggulan dan kelemahan masing-masing, sehingga terkadang sulit untuk menentukan satu model terbaik yang dapat menghasilkan prediksi yang paling akurat di semua kondisi (Yu et al., 2010). ML dapat menggabungkan berbagai algoritma untuk meningkatkan akurasi prediksi dan memberikan wawasan





tentang faktor risiko, sehingga mendukung strategi mitigasi yang lebih proaktif dan responsif.

Ensemble learning adalah solusi yang menggabungkan beberapa model ML untuk memperkuat hasil prediksi. *Ensemble learning* bekerja berdasarkan prinsip bahwa gabungan dari beberapa model yang berbeda dapat menghasilkan prediksi yang lebih baik daripada hanya mengandalkan satu model saja (Priasni dan Oswari, 2021). Salah satu variasi dari *ensemble learning* yang semakin populer adalah *Weighted Ensemble* (Zhou, 2015).

Pada penelitian ini, penulis akan mengembangkan dan mengimplementasikan model *Weighted Ensemble* untuk meramalkan pemadaman listrik yang disebabkan oleh cuaca ekstrem. Penelitian ini akan memanfaatkan berbagai indikator cuaca untuk membangun model prediksi yang akurat. Model diharapkan memberikan hasil prediksi yang lebih presisi dibandingkan model individu. Hasil penelitian ini diharapkan dapat membantu operator jaringan listrik dalam membuat langkah mitigasi yang lebih efektif, meningkatkan ketahanan infrastruktur listrik terhadap cuaca ekstrem, serta mengurangi risiko kerugian ekonomi dan keselamatan akibat pemadaman listrik.

B. Rumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan, rumusan masalah pada penelitian ini adalah sebagai berikut.

1. Bagaimana membangun metode prediksi menggunakan *Weighted Ensemble* untuk meningkatkan akurasi prediksi pemadaman listrik akibat cuaca ekstrem yang dibandingkan dengan model secara individu?
2. Bagaimana cara menangani *missing values* dalam data untuk memastikan model *Weighted Ensemble* dapat memberikan prediksi pemadaman listrik akibat cuaca ekstrem yang lebih akurat?

C. Tujuan

Berdasarkan rumusan masalah yang telah dipaparkan, tujuan pada penelitian ini adalah sebagai berikut.

1. Membangun metode prediksi menggunakan *Weighted Ensemble* untuk meningkatkan akurasi prediksi pemadaman listrik akibat cuaca ekstrem yang dibandingkan dengan model secara individu.
2. Mengimplementasikan teknik penanganan *missing values* dalam data untuk memastikan model *Weighted Ensemble* dapat memberikan prediksi pemadaman listrik akibat cuaca ekstrem yang lebih akurat.





BAB 2 METODOLOGI

A. Deskripsi *Dataset*

Dataset yang digunakan adalah data *time series* yang menyajikan perubahan nilai indikator-indikator cuaca terhadap waktu selama dua tahun, dari tahun 2022–2024. *Dataset* tersebut memiliki 34 kolom yang digunakan untuk meramalkan cuaca ekstrem dan waktu yang tepat untuk mematikan listrik. Tabel 2.1 menunjukkan indikator-indikator cuaca yang direpresentasikan sebagai fitur. Fitur tersebut akan digunakan dalam analisis.

Tabel 2.1 Keterangan *Dataset*

Fitur	Jenis Data	Deskripsi
temperature_2m (°C)	Numerik	Suhu udara pada ketinggian 2 meter dari permukaan tanah
relative_humidity_2m (%)	Numerik	Kelembapan relatif pada ketinggian 2 meter dari permukaan tanah
dew_point_2m (°C)	Numerik	Suhu titik embun pada ketinggian 2 meter
apparent_temperature (°C)	Numerik	Suhu yang dirasakan oleh tubuh manusia
pressure_msl (hPa)	Numerik	Tekanan atmosfer pada permukaan laut
surface_pressure (hPa)	Numerik	Tekanan atmosfer pada permukaan tanah
cloud_cover (%), cloud_cover_low (%), cloud_cover_mid (%), dan cloud_cover_high (%)	Numerik	Persentase langit yang tertutup awan pada waktu tertentu berdasarkan ketinggian
vapour_pressure_deficit (kPa)	Numerik	Selisih antara tekanan uap air maksimum dengan tekanan uap air aktual.
wind_speed_10m (km/h) dan wind_speed_100m (km/h)	Numerik	Kecepatan angin pada ketinggian tertentu dari permukaan tanah
wind_direction_10m (°) dan wind_direction_100m (°)	Numerik	Arah angin pada ketinggian tertentu dari permukaan tanah
wind_gusts_10m (km/h)	Numerik	Kecepatan hembusan angin
soil_temperature_0_to_7cm, soil_temperature_7_to_28cm , soil_temperature_28_to_100cm , dan soil_temperature_100_to_255c m (°C)	Numerik	Suhu tanah pada kedalaman tertentu dari permukaan





Fitur	Jenis Data	Deskripsi
soil_moisture_0_to_7cm, soil_moisture_7_to_28cm, soil_moisture_28_to_100cm, dan soil_moisture_100_to_255cm (m ³ /m ³)	Numerik	Kadar air tanah pada kedalaman tertentu
shortwave_radiation (W/m ²)	Numerik	Radiasi gelombang pendek yang mencapai permukaan tanah
direct_radiation (W/m ²)	Numerik	Radiasi langsung dari matahari yang mencapai permukaan tanah
diffuse_radiation (W/m ²)	Numerik	Radiasi tersebar yang mencapai permukaan tanah
direct_normal_irradiance (W/m ²)	Numerik	Radiasi langsung yang diukur tegak lurus terhadap permukaan tanah
global_tilted_irradiance (W/m ²)	Numerik	Radiasi total yang diterima oleh permukaan miring
terrestrial_radiation (W/m ²)	Numerik	Radiasi yang dipancarkan oleh permukaan bumi
is_day	Kategorik	Indikator apakah waktu pengambilan data adalah siang hari (1) atau malam hari (0)
sunshine_duration (s)	Numerik	Durasi sinar matahari yang tercatat selama waktu tertentu
precipitation (mm)	Numerik	Jumlah presipitasi (hujan) yang tercatat selama waktu tertentu
electricity_shutdown	Kategorik	Indikator apakah terjadi pemadaman listrik pada waktu tertentu (1) atau tidak (0)

Nilai tiap fitur direkam setiap 1 jam sehingga *dataset* memiliki 21.888 baris. Hal tersebut memberikan gambaran yang komprehensif terkait perubahan indikator cuaca yang terbagi menjadi tiga bagian, yakni bagian bawah tanah, permukaan tanah, dan bagian atas tanah. Data terbagi menjadi data latih dan data uji. Data latih digunakan untuk melatih model agar mendapatkan parameter yang terbaik dalam memprediksi dua fitur target. Fitur target yang pertama adalah curah hujan yang dapat digunakan bersama fitur sebelumnya untuk memprediksi fitur target kedua, yaitu 'electricity_shutdown'. Parameter yang telah didapatkan melalui pelatihan dapat digunakan untuk meramalkan fitur target di beberapa jam kedepan dengan data uji yang ada.





B. Metode Analisis

1. Analisis Data Eksploratori

Analisis Data Eksploratori (ADE) adalah proses pertama yang dilakukan setelah menyiapkan data. Proses ADE dilakukan di awal karena termasuk langkah penting untuk mendapatkan wawasan dan informasi dari data dengan cara yang beragam (Smith-Miles 2012). Umumnya gambaran tersebut disajikan melalui tampilan visual agar dapat lebih dipahami. Salah dua gambaran statistik yang dapat dianalisis untuk memahami pesan yang terkandung dalam data adalah distribusi dan korelasi tiap fitur dengan fitur target.

a. Distribusi Tiap Fitur

Ketidakpastian suatu data dapat diekspresikan melalui distribusi. Distribusi menyajikan sebaran data dengan menunjukkan frekuensi munculnya datum yang korespon dengan fitur tertentu (Mishra & Datta-Gupta 2018). Sebaran tersebut dapat menunjukkan informasi statistik yang dapat berguna untuk keperluan analisis, seperti keberagaman dan nilai pusat dari data (Wild 2006). Salah satu distribusi yang menjadi dasar bagi banyak distribusi lainnya adalah distribusi normal. Distribusi normal sangat sesuai dengan *feature scaling* yang dikenal sebagai *standard scaler*, sementara data yang tidak berdistribusi normal lebih cocok menggunakan *min-max scaler* (Chou dkk. 1991; Bosman 2007).

b. Mutual Information (MI)

Korelasi digunakan untuk mengukur hubungan antara setiap fitur indikator cuaca terhadap fitur curah hujan dan *'electricity_shutdown'*. Korelasi tersebut dapat diukur menggunakan metode *mutual information*. MI adalah metode untuk mengukur dependensi antar dua variabel dengan menghitung tingkat pengetahuan satu variabel dalam mereduksi ketidakpastian variabel yang lain (Newman, Cantwell & Young 2019).

MI antara dua variabel, yaitu fitur X dan fitur target Y, diperoleh dengan menghitung jumlah dari peluang gabungan $p(x, y)$ dikalikan dengan logaritma dari perbandingan antara peluang gabungan dan perkalian peluang marginal $p(x)$ dan $p(y)$ dari masing-masing fitur (Kwak 2002). Secara matematis, MI dinyatakan sebagai

$$I(X; Y) = \int p(x, y) \log\left(\frac{p(x, y)}{p(x)p(y)}\right) dx dy.$$

2. Data Preprocessing

Dataset yang diberikan tidak semata-mata bersih, seperti perbedaan skala antar fitur, nilai kosong pada beberapa fitur, dan kemungkinan adanya fitur lain yang memiliki informasi penting. Oleh karena itu, langkah *data preprocessing* diperlukan untuk membersihkan data tersebut. *Data preprocessing* adalah proses untuk mengurangi kompleksitas *dataset*, sehingga *dataset* yang mentah menjadi lebih mudah





dipahami dan memungkinkan interpretasi yang baik oleh model (Ramírez-Gallego dkk. 2017).

a. *Feature Scaling*

Feature scaling atau yang dikenal dengan normalisasi adalah teknik mentransformasi nilai-nilai dari suatu fitur agar memiliki skala atau rentang nilai yang umum (Wan 2019). Dengan kata lain, jangkauan nilai dari setiap fitur tidak berbeda jauh, sehingga dapat memudahkan komputasi fungsi kesalahan. Hal tersebut akan sangat berpengaruh pada performa model (TharaD, PremaSudhaB, & Xiong 2019). Minimasi fungsi kesalahan akan lebih mudah mencapai konvergensi sehingga dapat menghemat waktu komputasi.

Normalisasi yang digunakan adalah *min-max scaler* karena distribusi setiap fitur pada *dataset* tidak berdistribusi normal. *Min-max* adalah teknik normalisasi yang mengubah nilai suatu fitur ke rentang tertentu, umumnya 0 hingga 1. Nilai suatu fitur diubah menjadi perbandingan antara rentang nilai fitur X dan nilai minimum X_{\min} dari fitur tersebut dengan nilai maksimum X_{\max} dan X_{\min}

$$X_{\text{saled}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}}.$$

b. *Feature Engineering*

Terkadang fitur yang ada dari *dataset* yang mentah belum memberikan informasi yang cukup relevan dengan fitur target yang diinginkan. Permasalahan tersebut dapat diselesaikan dengan *feature engineering*. *Feature engineering* adalah teknik menentukan fitur yang relevan, sehingga memberikan informasi yang baik pada model.

- *Konversi fitur waktu*, mengubah fitur 'time' menjadi format *datetime* dan mengekstraknya menjadi fitur tahun, bulan, hari, dan jam.

$$\sin \text{ waktu} = \sin\left(\frac{\text{waktu} \cdot 2\pi}{\text{periode}}\right), \cos \text{ waktu} = \cos\left(\frac{\text{waktu} \cdot 2\pi}{\text{periode}}\right).$$

- *Mengubah fitur angin menjadi vektor*, kecepatan dan arah angin dari representasi polar diubah menjadi representasi kartesian dengan mengalikan nilai kecepatan atau arah angin v dan fungsi trigonometri \sin dan \cos untuk mendapatkan dua komponen baru pada sumbu x , v_x , dan sumbu y , v_y . Secara matematis dapat dinyatakan sebagai berikut

$$\bar{v}_x = v \cdot \cos\left(\theta \frac{\pi}{180}\right) \text{ dan } \bar{v}_y = v \cdot \sin\left(\theta \frac{\pi}{180}\right).$$

- *Menambah rate of change (ROC)*, fitur ini menghitung perubahan nilai antarwaktu untuk setiap indikator lingkungan. ROC didapat dengan menghitung perbandingan selisih nilai antara setiap baris dan baris sebelumnya dengan selisih waktu. ROC dalam *time series* biasanya diukur dengan menghitung perubahan persentase dari satu periode waktu ke periode berikutnya.

$$ROC = \frac{\text{nilai sekarang} - \text{nilai sebelumnya}}{\text{nilai sebelumnya}} 100.$$



- *Menambah fitur lag*, fitur ini dapat diperoleh menggunakan nilai-nilai yang berada pada titik waktu sebelumnya dengan melakukan *shifting*. Fitur ini dapat memberikan informasi temporal yang membantu model untuk menemukan pola perubahan data dari waktu ke waktu.

c. Pengisian Nilai Kosong

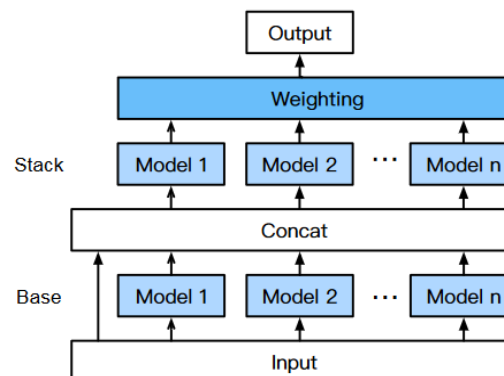
Dataset yang tersedia tidak selalu lengkap akibat berbagai kemungkinan dalam proses pengumpulan data. Oleh karena itu, perlakuan tertentu perlu dilakukan untuk mengisi nilai kosong agar integritas data terjaga dan informasi yang dihasilkan tetap berkualitas untuk model. Berbagai cara dapat digunakan untuk mengisi nilai kosong, diantaranya adalah mengisi menggunakan model *K-Nearest Neighbours Imputer* (KNNI) dan *Weighted Ensemble*. Pengisian nilai kosong dimulai dari fitur yang memiliki *mutual information* terendah dengan fitur target menggunakan KNNI. Langkah ini diambil untuk menghemat waktu komputasi. Sementara itu, nilai kosong pada fitur dengan MI yang lebih tinggi akan diisi menggunakan model *Weighted Ensemble*.



Gambar 2.1. Alur pengisian nilai kosong

3. Modelling

Tahap yang sangat penting dalam prediksi fitur target adalah *modelling*. Model yang digunakan untuk memprediksi kedua fitur target adalah *Weighted Ensemble*. Ide utama dari model *Weighted Ensemble* adalah mengurangi *loss* melalui agregasi hasil prediksi dari beberapa model dasar yang berbeda, Gambar 2.2 (Li dkk 2016). Hasil prediksi atau *loss* yang diperoleh dari masing-masing model kemudian diberi bobot dengan harapan agar setiap model dapat saling mengimbangi *loss* antara satu model satu dengan yang lain, sehingga dapat menghasilkan model yang mereduksi *loss* tersebut (Zhang & Zhou 2011).



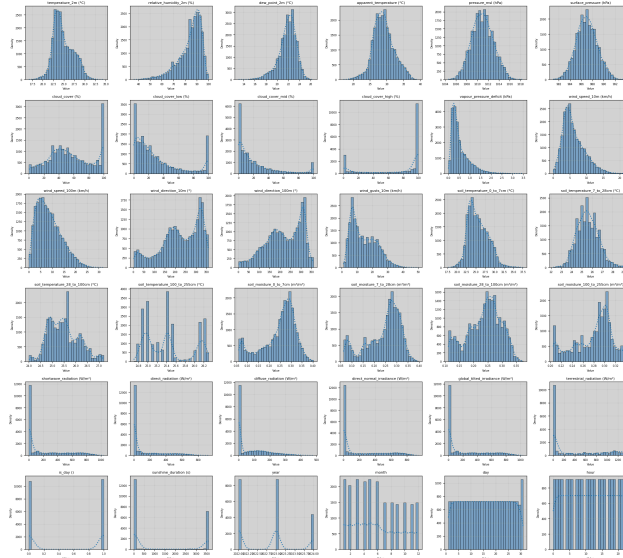
Gambar 2.2. Strategi *Weighted Ensemble*

BAB 3 HASIL DAN PEMBAHASAN

A. Analisis

Pada tahap analisis, Analisis Data Eksploratori (ADE) akan dimulai dengan menggabungkan data *train* dan *test*. Nilai kosong akan diisi menggunakan *KNN imputer*, sementara kolom 'precipitation' diisi dengan nilai rata-rata (*mean*).

1. Analisis distribusi akan dilakukan terhadap semua fitur yang terdapat dalam tabel.

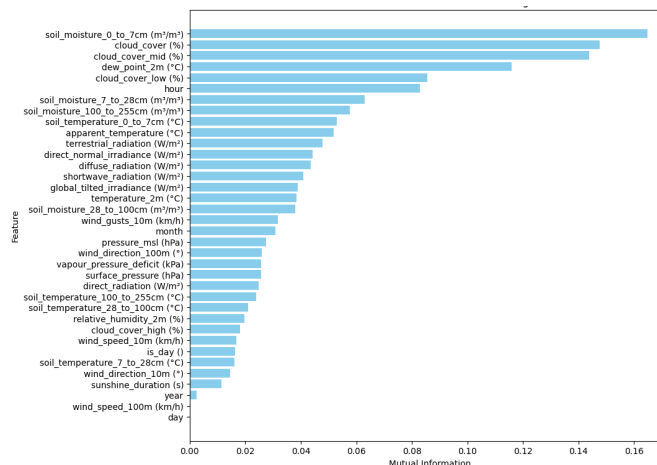


Gambar 3.1. Distribusi untuk setiap fitur

Berdasarkan Gambar 3.1, mayoritas fitur memiliki distribusi yang *skewed* dan *multimodal*, bukan distribusi normal. Distribusi *skewed* menunjukkan bahwa data tidak simetris, sementara *multimodal* berarti terdapat lebih dari satu puncak, mencerminkan adanya sub populasi dalam data. Fitur-fitur ini tidak mengikuti distribusi normal dan pola yang bervariasi, sehingga pemilihan metode normalisasi dengan *min-max scaler* yang mengubah data ke rentang 0 hingga 1 sangat tepat. Metode ini efektif untuk menjaga konsistensi skala fitur, terutama bagi model yang sensitif terhadap perbedaan skala.

2. Analisis MI antara fitur dengan target

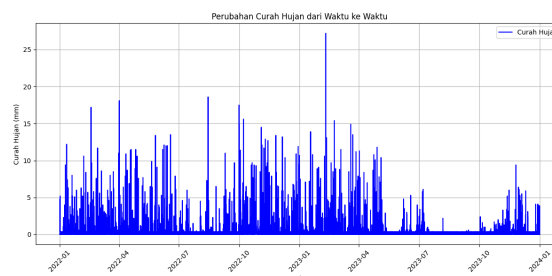
MI setiap fitur dengan target akan dianalisis untuk mengukur ketergantungan antar fitur dan seberapa besar kontribusi masing-masing fitur dalam menjelaskan variabilitas target. Analisis ini juga membantu mengidentifikasi keterkaitan atau prediktabilitas antar fitur.



Gambar 3.2. MI

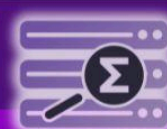
Dari gambar di atas, terlihat bahwa fitur “soil_moisture_0_to_7cm”, “cloud_cover”, “dew_point_2m”, dan “soil_moisture_7_to_28cm” memiliki nilai MI tinggi, menunjukkan keterikatan kuat dengan target. Sebaliknya, fitur seperti “sunshine_duration”, “wind_speed_100m”, dan “wind_direction_10m” memiliki nilai MI rendah yang menunjukkan kontribusi kecil. Fitur dengan MI rendah dan nilai kosong akan diisi menggunakan *KNN imputer* karena pengaruhnya kecil. Sementara itu, fitur dengan MI tinggi akan diisi menggunakan model kompleks *Weighted Ensemble* untuk memastikan kualitas informasi yang akurat.

3. Analisis perubahan curah hujan dari waktu ke waktu



Gambar 3.3. Perubahan curah hujan

Gambar 3.3 menunjukkan perubahan curah hujan dalam milimeter (mm) dari waktu ke waktu, dengan sumbu X sebagai waktu dan sumbu Y sebagai curah hujan. Terlihat bahwa curah hujan berfluktuasi cukup besar sepanjang periode yang diamati, terutama di awal tahun 2023 yang terjadi lonjakan curah hujan cukup signifikan. Fluktuasi curah hujan yang cukup mencolok juga muncul di pertengahan 2022 dan akhir 2023. Grafik ini memberikan gambaran jelas tentang variasi curah hujan yang terjadi selama periode tersebut.



4. Analisis perubahan kecepatan angin dan arah angin dari waktu ke waktu



Gambar 3.4. Perubahan kecepatan dan suhu angin

Gambar 3.4 menampilkan dua grafik garis yang menggambarkan perubahan kecepatan dan arah angin pada ketinggian 10 meter dan 100 meter seiring waktu. Grafik pertama menunjukkan bahwa angin di 100 meter lebih kuat, sering mencapai lebih dari 20 km/h, terutama dari pertengahan Februari hingga awal Maret, dengan fluktuasi signifikan. Grafik kedua memperlihatkan arah angin yang konsisten meskipun ada variasi, dan hembusan angin (*wind gusts*) pada 10 meter dengan fluktuasi lebih rendah dibandingkan kecepatan angin. Kedua grafik ini mengilustrasikan fluktuasi kecepatan dan arah angin yang kompleks selama periode observasi.

B. Pembahasan

1. *Data Preprocessing*

Pada tahap *data preprocessing*, akan dibahas mengenai pengisian nilai yang hilang (*missing values*) serta penambahan fitur dalam *dataset*. Untuk mempermudah proses penanganan *missing values*, data *train* dan *test* akan digabungkan terlebih dahulu sebelum dilakukan *preprocessing* lebih lanjut.

a. Normalisasi

Berdasarkan hasil analisis ADE sebelumnya, proses normalisasi menggunakan metode *min-max* akan diterapkan. Hal ini dilakukan karena banyak distribusi data yang tidak mengikuti distribusi normal, sehingga metode *min-max* lebih sesuai untuk meratakan skala data tanpa mengubah struktur distribusinya. Normalisasi ini juga penting untuk menjaga rentang nilai fitur tetap konsisten, terutama ketika ada perbedaan yang



signifikan antar skala data, sehingga model pembelajaran mesin dapat bekerja lebih optimal dan akurat.

b. Missing values

Dataset memiliki nilai yang hilang (*missing values*) yang dapat dilihat pada Tabel 3.1 di bawah ini.

Tabel 3.1. *Missing values* pada *dataset*

Nama Fitur	Banyak Nilai Kosong	Persentase Nilai Kosong (%)
temperature_2m (°C)	9988	45.632310
cloud_cover (%)	5461	24.949744
cloud_cover_low (%)	4449	20.326206
wind_speed_10m (km/h)	4384	20.029240
wind_gusts_10m (km/h)	3273	14.953399
cloud_cover_mid (%)	3250	14.848319
surface_pressure (hPa)	2264	10.343567
wind_direction_100m (°)	2246	10.261330
relative_humidity_2m (%)	2206	10.078582
vapour_pressure_deficit (kPa)	2205	10.074013
soil_moisture_28_to_100cm (m ³ /m ³)	2192	10.014620
sunshine_duration (s)	2153	9.836440
wind_direction_10m (°)	2138	9.767909
terrestrial_radiation (W/m ²)	2119	9.681104

Selain itu, terdapat nilai kosong pada label di data *train* yang label *precipitation* memiliki 3.542 nilai kosong, dan label *electricity shutdown* memiliki 3.531 nilai kosong. Berdasarkan analisis MI yang telah dijelaskan pada ADE, kolom yang akan diisi menggunakan KNNI dengan 10 *neighbours* adalah 'wind_speed_10m (km/h)', 'wind_gusts_10m (km/h)', 'surface_pressure (hPa)', 'wind_direction_100m (°)', 'relative_humidity_2m (%)', dan 'vapour_pressure_deficit (kPa)'. Kolom-kolom ini dipilih karena memiliki nilai MI yang relatif rendah terhadap target, sehingga penggunaan *KNN imputer* dianggap memadai untuk mengisi nilai kosong tanpa berdampak signifikan pada hasil model.

Selanjutnya, setiap nilai kosong yang tersisa pada fitur lainnya akan diisi menggunakan model terbaik yang dihasilkan dari proses *stacking* model-model *machine learning* tradisional dan *deep learning*, yang digabungkan menjadi model *Weighted Ensemble*, sebagaimana ditunjukkan pada tabel di bawah ini.

Tabel 3.2. Model pengisian nilai kosong

Nama Fitur	Model	Metriks	Skor Validasi
temperature_2m (°C)	WeightedEnsemble_L2	RMSE	0.0056
cloud_cover (%)	WeightedEnsemble_L2	RMSE	0.079
cloud_cover_low (%)	WeightedEnsemble_L2	RMSE	0.092





Nama Fitur	Model	Metriks	Skor Validasi
cloud_cover_mid (%)	WeightedEnsemble_L2	RMSE	1.3722
soil_moisture_28_to_100cm (m ³ /m ³)	WeightedEnsemble_L2	RMSE	0.0055
sunshine_duration (s)	WeightedEnsemble_L2	RMSE	0.0025
wind_direction_10m (°)	WeightedEnsemble_L2	RMSE	0.1533
terrestrial_radiation (W/m ²)	WeightedEnsemble_L2	RMSE	0.0054
precipitation (mm)	WeightedEnsemble_L2	RMSE	0.2359

Pengisian nilai kosong dilakukan menggunakan model WeightedEnsemble_L2, dengan waktu komputasi sekitar 10 menit per fitur, yang tergolong sangat cepat. Selain efisien, metode ini juga memberikan performa yang baik, sehingga mampu mengisi nilai kosong dengan akurasi optimal. Namun, pada label *electricity shutdown*, pengisian belum dilakukan dan akan dianalisis lebih lanjut karena adanya ketidakseimbangan data. Oleh karena itu, dalam proses selanjutnya, label *electricity shutdown* akan di-drop sementara waktu hingga analisis lebih mendalam dapat dilakukan.

c. Ekstraksi fitur angin menjadi vektor

Pada data, terdapat lima fitur yang terkait dengan angin, yaitu *windspeed 10m (km/h)*, *windgusts 10m (km/h)*, *windspeed 100m (km/h)*, *winddirection 10m (°)*, dan *winddirection 100m (°)*. Fitur-fitur angin ini sebaiknya tidak dimasukkan secara langsung ke dalam model karena fitur *winddirection* bersifat siklis. Sebagai contoh, jarak antara 0° dan 359° seharusnya lebih dekat daripada 0° dan 10°, tetapi model tidak mampu menangkap sifat siklis ini secara alami.

Pembuatan fitur interaksi yang merupakan kombinasi antara *windspeed* dan *windgusts* dengan *winddirection* diperlukan, sehingga lebih baik dalam menjelaskan kecepatan dan arah angin. Dengan transformasi ini, distribusi dari fitur angin menjadi lebih terpusat. Perbandingan antara distribusi fitur angin tanpa rekayasa fitur dan setelah dipecah menjadi vektor sumbu x dan y menunjukkan bahwa transformasi ini dapat memberikan representasi yang lebih baik mengenai kecepatan dan arah angin dalam model.

d. Penambahan fitur *Rate Of Change* (ROC)

Berdasarkan hasil ADE di atas, terlihat perubahan pada curah hujan, suhu, kelembapan tanah, serta kecepatan dan arah angin yang terjadi setiap hari, bahkan setiap jam. Oleh karena itu, ROC menjadi sangat penting untuk memahami pola yang mungkin mengalami fluktuasi signifikan atau tetap stabil tergantung waktu. Rumus ini memberikan persentase perubahan antar periode, yang berguna untuk menganalisis pola perubahan yang dinamis, seperti deteksi tren naik, turun, atau stabilitas variabel dalam konteks waktu. ROC membantu dalam mengidentifikasi pergerakan signifikan dalam variabel cuaca dan lingkungan, sehingga menjadi





indikator penting dalam analisis data *time series*. Tabel 3.3 adalah contoh setelah ROC.

Tabel 3.3. Contoh Tampilan Setelah dilakukan ROC

temperature_2m (°C)	roc_temperature_2m (°C)
0.31817471981048584	
0.3089887640449438	-0.009185955765542042
0.329636812210083	0.02064804816513921

e. Penambahan fitur *lag*

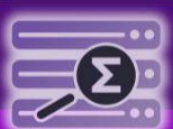
Waktu-waktu di masa lalu dapat berpengaruh terhadap fitur-fitur di masa depan. Sebagai contoh, *cloud cover* (tutupan awan) lima jam yang lalu mungkin memengaruhi kondisi cuaca beberapa jam kemudian. Begitu pula dengan arah angin yang terjadi lima jam sebelumnya bisa berkontribusi terhadap perubahan *cloud cover* di masa depan. Oleh karena itu, dalam analisis ini ditambahkan fitur *lag* dengan *5 lag steps*. Tujuannya adalah untuk menangkap keterkaitan antara kondisi cuaca di masa lalu dengan keadaan cuaca di masa depan, sehingga model dapat lebih baik dalam memprediksi pola cuaca dengan mempertimbangkan faktor-faktor historis yang relevan. Tabel 3.4 adalah contoh setelah dilakukan *lag*.

Tabel 3.4. Contoh Tampilan Setelah dilakukan *lag* dengan *steps 5*

temperature_2m (°C)	temperature_2m (°C)_Lag1	temperature_2m (°C)_Lag2	temperature_2m (°C)_Lag3	temperature_2m (°C)_Lag4	temperature_2m (°C)_Lag5
0.318174					
0.308988	0.318174				
0.329636	0.308988	0.318174			
0.325842	0.329636	0.308988	0.318174		
0.337078	0.325842	0.329636	0.308988	0.318174	
0.3202247	0.337078	0.325842	0.329636	0.308988	0.318174

2. Modeling

Kolom target dalam penelitian ini adalah *precipitation* dan *electricity_shutdown* yang akan diprediksi menggunakan model *Weighted Ensemble*. Semua model yang digunakan untuk membentuk *Weighted Ensemble* dilatih baik pada lapisan pertama (*first stack*) maupun lapisan kedua (*second stack*). Namun, perlu diperhatikan bahwa kolom label *electricity_shutdown* masih mengandung *missing values*, sehingga strategi yang diterapkan adalah menghapus baris yang memiliki nilai kosong tersebut sebelum melatih model untuk target ini.





Langkah-langkah yang dilakukan adalah sebagai berikut: pertama, *dataset* dilatih untuk memprediksi precipitation tanpa melibatkan label *electricity_shutdown*. Setelah prediksi precipitation pada data uji selesai, baris-baris dengan nilai kosong pada *electricity_shutdown* dihapus karena pengisian nilai yang hilang dengan model seperti pada precipitation tidak memungkinkan. Hal ini disebabkan oleh ketidakseimbangan data pada label *electricity_shutdown* ketika jumlah nilai 1 jauh lebih sedikit dibandingkan nilai 0.

Setelah proses pembersihan data selesai, model dilatih untuk memprediksi *electricity_shutdown* menggunakan *dataset* yang telah dibersihkan dari nilai kosong. Hasil dari prediksi ini menghasilkan dua papan peringkat model: satu untuk precipitation pada Tabel 3.5 dan satu lagi untuk *electricity_shutdown* pada Tabel 3.6 yang diurutkan berdasarkan skor validasi dari masing-masing model.

Tabel 3.5. Papan peringkat model (*precipitation*)

Model	Skor Validasi (RMSE)	Waktu Pelatihan Model (detik)
WeightedEnsemble_L2	0.1262	0.03
CatBoost	0.1518	442.54
LightGBMXT	0.1769	123.66
NeuralNetFastAI	0.1968	27.8
LightGBM	0.1947	28.86
ExtraTreesMSE	0.226	88.25
RandomForestMSE	0.2541	557.86
XGBoost	0.2399	38.49

Tabel 3.5 menunjukkan peringkat model dalam prediksi precipitation berdasarkan skor validasi RMSE dan waktu pelatihan. Model terbaik adalah WeightedEnsemble_L2 dengan RMSE 0.1262 dan waktu pelatihan 0.03 detik. CatBoost berada di posisi kedua (RMSE 0.1518, waktu 442.54 detik). Model lain seperti LightGBMXT (RMSE 0.1769), NeuralNetFastAI (RMSE 0.1968), dan LightGBM (RMSE 0.1947) juga menunjukkan kinerja baik, sementara ExtraTreesMSE, RandomForestMSE, dan XGBoost memiliki RMSE lebih tinggi dan waktu pelatihan lebih lama.

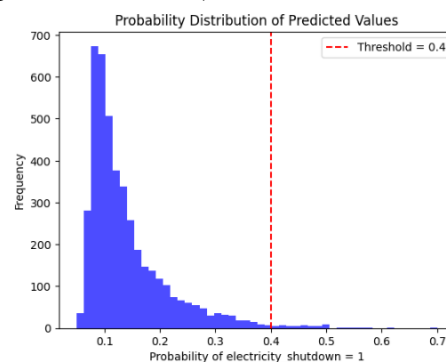
Tabel 3.6. Papan peringkat model (*electricity_shutdown*)

Model	Skor Validasi (<i>F1 Score</i>)	Waktu Pelatihan Model (detik)
WeightedEnsemble_L2	0.2857	0.03
CatBoost	0.1333	26.81
LightGBMXT	0.2424	9.69
NeuralNetFastAI	0.0115	16.1
LightGBM	0.1481	11.56
ExtraTreesGini	0.0	2.5
RandomForestGini	0.0	7.3
XGBoost	0.1765	8.39



Tabel 3.6 menyajikan peringkat model dalam memprediksi *electricity_shutdown* berdasarkan *F1 Score* dan waktu pelatihan. Model terbaik adalah *WeightedEnsemble_L2* dengan *F1 Score* 0.2857 dan waktu pelatihan 0.03 detik, diikuti oleh *LightGBMXT* (*F1 Score* 0.2424) dan *XGBoost* (*F1 Score* 0.1765). Model lainnya seperti *CatBoost* dan *LightGBM* memiliki skor yang lebih rendah, sementara *ExtraTreesGini* dan *RandomForestGini* mencatat *F1 Score* 0.0.

Namun, *F1 Score* secara keseluruhan relatif rendah yang menunjukkan tantangan dalam memprediksi target dengan akurasi yang baik. Oleh karena itu, *threshold* prediksi binary yang biasanya berada pada 50% diturunkan menjadi 40% untuk meningkatkan sensitivitas model terhadap kelas '1' (*shutdown*), agar lebih terlihat dalam prediksi. Hasil dari penurunan *threshold* ini dapat dilihat pada gambar berikut (disertakan visualisasi).



Gambar 3.5. Distribusi probabilitas prediksi

Gambar 3.5 menunjukkan distribusi probabilitas prediksi untuk label *electricity_shutdown* dengan *threshold* sebesar 0.4. Pada sumbu horizontal (*x-axis*) adalah probabilitas *shutdown* = 1, dan sumbu vertikal (*y-axis*) menunjukkan frekuensi prediksi dengan probabilitas tertentu.

Mayoritas prediksi memiliki probabilitas yang sangat rendah, sebagian besar berada di bawah 0.1, menunjukkan bahwa sebagian besar model cenderung memprediksi "shutdown" tidak terjadi (label 0). Namun, dengan menurunkan *threshold* menjadi 0.4 (garis merah putus-putus), model akan lebih sering memprediksi label 1 (*shutdown*), bahkan ketika probabilitasnya tidak mencapai 0.5, meningkatkan sensitivitas terhadap kelas positif (*shutdown*).

Dengan *threshold* 0.4, meskipun prediksi label 1 lebih sering terjadi, hal ini juga berpotensi meningkatkan prediksi positif palsu, sehingga perlu keseimbangan antara sensitivitas dan presisi tergantung pada tujuan prediksi.



BAB 4

KESIMPULAN

A. Kesimpulan

Penelitian ini bertujuan untuk mengembangkan sistem peramalan berbasis indikator cuaca yang dapat memprediksi dampak cuaca ekstrem terhadap jaringan listrik. Sistem ini dirancang menggunakan Python sebagai perangkat utama, dengan menerapkan berbagai teknik analisis data dan machine learning. Salah satu inovasi dalam penelitian ini adalah penggunaan feature extraction yang mencakup penambahan fitur seperti rate of change dan lag untuk menangkap pola dinamika cuaca di masa lalu yang mempengaruhi peristiwa di masa depan. Teknik ini terbukti efektif dalam meningkatkan akurasi prediksi model.

Model yang dikembangkan berhasil mengidentifikasi indikator cuaca yang signifikan, seperti kecepatan angin, arah angin, dan curah hujan, serta memodelkan hubungan antara indikator tersebut dengan frekuensi dan lokasi pemadaman listrik. Evaluasi model dilakukan dengan menggunakan dua tugas prediksi: regresi untuk memprediksi curah hujan (precipitation) dan klasifikasi untuk memprediksi pemadaman listrik (electricity_shutdown). Hasil evaluasi menunjukkan bahwa pada tugas regresi, model Weighted Ensemble L2 mencapai skor Root Mean Square Error (RMSE) sebesar 0.1262, sementara pada tugas klasifikasi, model yang sama mencapai skor F1-Score sebesar 0.2857.

Meskipun skor F1 Score relatif rendah, penyesuaian pada threshold klasifikasi dari 0.5 menjadi 0.4 berhasil meningkatkan sensitivitas terhadap prediksi label shutdown (kelas 1). Penurunan threshold ini memungkinkan model untuk lebih akurat dalam mendeteksi kemungkinan pemadaman listrik, meskipun terdapat peningkatan risiko prediksi positif palsu.

Secara keseluruhan, sistem ini menunjukkan potensi yang signifikan dalam meningkatkan ketahanan infrastruktur jaringan listrik terhadap cuaca ekstrem dan bencana meteorologi. Hasil penelitian ini dapat berfungsi sebagai alat yang berguna bagi operator jaringan listrik untuk merencanakan tindakan mitigasi yang lebih efektif dan memperkuat sistem distribusi tenaga listrik. Dengan peramalan yang lebih akurat, risiko gangguan operasional akibat cuaca ekstrem dapat diminimalkan, memberikan manfaat besar bagi keberlanjutan sistem kelistrikan di masa depan.





DAFTAR PUSTAKA

- Clarke, B., Otto, F.E., Stuart-Smith, R.F. & Harrington, L.J 2022, 'Extreme weather impacts of climate change: an attribution perspective', *Environmental Research: Climate*, 1.
- Fawzy, S., Osman, A., Doran, W. & Rooney, D 2020, 'Strategies for mitigation of climate change: a review', *Environmental Chemistry Letters*, 18, pp. 1-16. DOI: 10.1007/s10311-020-01059-w.
- Añel, J., Pérez-Souto, C., Bayo-Besteiro, S., Prieto-Godino, L., Bloomfield, H., Troccoli, A. & Torre, L 2024, 'Extreme weather events and the energy sector in 2021', *Weather, Climate, and Society*, 16. DOI: 10.1175/WCAS-D-23-0115.1.
- Bollinger, L. and Dijkema, G 2016, 'Evaluating infrastructure resilience to extreme weather – the case of the Dutch electricity transmission network', *European Journal of Transport and Infrastructure Research*, 16(1), pp. 1-20. DOI: 10.18757/ejtr.2016.16.1.3122.
- Panteli, M. and Mancarella, P 2015, 'Influence of extreme weather and climate change on the resilience of power systems: impacts and possible mitigation strategies', *Electric Power Systems Research*, 127, pp. 259-270. DOI: 10.1016/j.epsr.2015.06.012.
- Yu, L., Wang, S. and Lai, K 2010, 'Developing an SVM-based ensemble learning system for customer risk identification collaborating with customer relationship management', *Frontiers of Computer Science in China*, 4(2), pp. 196-203. DOI: 10.1007/s11704-010-0508-2.
- Priasni, T. and Oswari, T 2021, 'Comparative study of standalone classifier and ensemble classifier', *Telkomnika (Telecommunication Computing Electronics and Control)*, 19(5), p. 1747. DOI: 10.12928/telkomnika.v19i5.19508.
- Zhou, Z 2015, *Ensemble learning*, in *Encyclopedia of Machine Learning and Data Mining*, Springer, pp. 411-416. DOI: 10.1007/978-1-4899-7488-4_293.
- Smith-Miles, K. 2012, 'Exploratory data analysis', *Wiley Interdisciplinary Reviews: Computational Statistics*, 1. <https://doi.org/10.1002/wics.2>.
- Gelman, A. 2004, 'Exploratory Data Analysis for Complex Models', *Journal of Computational and Graphical Statistics*, 13, pp. 755 - 779. <https://doi.org/10.1198/106186004X11435>.
- Mishra, S., & Datta-Gupta, A 2018, *Chapter 3 – Distributions and Models Thereof*, pp. 31-67. <https://doi.org/10.1016/B978-0-12-803279-4.00003-1>.





- Wild, C 2006, 'The concept of distribution', *Statistics Education Research Journal*, 5, pp. 10-26. <https://doi.org/10.52041/serj.v5i2.497>.
- Newman, M., Cantwell, G., & Young, J 2019, 'Improved mutual information measure for classification and community detection', *Physical review. E*, 101 4-1, pp. 042304 . <https://doi.org/10.1103/PhysRevE.101.042304>.
- Kwak, N., & Choi, C 2002, 'Input Feature Selection by Mutual Information Based on Parzen Window', *IEEE Trans. Pattern Anal. Mach. Intell.*, 24, pp. 1667-1671. <https://doi.org/10.1109/TPAMI.2002.1114861>.
- Ramírez-Gallego, S., Krawczyk, B., García, S., Woźniak, M., & Herrera, F 2017, 'A survey on data preprocessing for data stream mining: Current status and future directions', *Neurocomputing*, 239, pp. 39-57. <https://doi.org/10.1016/J.NEUCOM.2017.01.078>.
- Chou, C., Bentler, P., & Satorra, A 1991, 'Scaled test statistics and robust standard errors for non-normal data in covariance structure analysis: a Monte Carlo study', *The British journal of mathematical and statistical psychology*, 44 (Pt 2), pp. 347-57 . <https://doi.org/10.1111/J.2044-8317.1991.TB00966.X>.
- Bosman, P., Grahl, J., & Rothlauf, F 2007, *SDR: a better trigger for adaptive variance scaling in normal EDAs*, pp. 492-499. <https://doi.org/10.1145/1276958.1277066>.
- Wan, X. 2019. 'Influence of feature scaling on convergence of gradient iterative algorithm', *Journal of Physics: Conference Series*, 1213. <https://doi.org/10.1088/1742-6596/1213/3/032021>.
- TharaD., K., PremaSudhaB., G., & Xiong, F 2019, 'Auto-detection of epileptic seizure events using deep neural network with different feature scaling techniques', *Pattern Recognit. Lett.*, 128, pp. 544-550. <https://doi.org/10.1016/j.patrec.2019.10.029>.
- Li, L., Hu, Q., Wu, X., & Yu, D, 2019, 'Exploration of classification confidence in ensemble learning', *Pattern Recognit.*, 47, 3120-3131. <https://doi.org/10.1016/j.patcog.2014.03.021>.
- Zhang, L., & Zhou, W, 2011. 'Sparse ensembles using weighted combination methods based on linear programming'. *Pattern Recognit.*, 44, 97-106. <https://doi.org/10.1016/j.patcog.2010.07.021>.

