BAB II
LANDASAN TEORI

2.1 Tinjauan Pustaka

2.1.1 *Random Forest Regression* Dalam *Forecasting* Beban Listrik

Berbagai studi terkini telah menunjukkan beberapa keunggulan *Random Forest Regression* dalam melakukan *forecasting* beban listrik jangka pendek baik dari akurasi, fleksibilitas, maupun efisiensi operasional. Penelitian oleh Magalhães, dkk. (2024) menyoroti bahwa diperlukan sebuah model yang mampu menangkap dinamika dari kompleksitas data beban listrik untuk melakukan *forecasting* jangka pendek. Hal ini disebabkan oleh tren non-linear, multipel musiman, dan variansi fluktuatif yang ada dalam data *time series* beban listrik itu sendiri. Dalam studinya, *Random Forest Regression* digunakan untuk memodelkan beban listrik jangka pendek menggunakan data historis dari empat gardu transformator di Universitas Beira Interior, Portugal. Studi ini menghasilkan tingkat akurasi yang tinggi dalam melakukan *forecasting* beban listrik jangka pendek melalui penyetelan (*tuning*) *hyperparameter* seperti jumlah pohon, jumlah minimum daun, dan jumlah fitur maksimal yang digunakan oleh model serta pendekatan teknik *feature engineering*. Lebih lanjut, penelitian ini juga menegaskan kontribusi *Random Forest Regression* terhadap pengambilan keputusan dalam *smart grid* serta efisiensi alokasi sumber daya dan pengurangan biaya operasional.

Studi oleh Vasenin, dkk. (2024) memperkuat argumen tersebut dengan membangun model *ensemble* yang menggabungkan *Random Forest Regression* dan *Histogram-based Gradient Boosting Regression* (HGBR). Hasil eksperimen pada data nyata dari fasilitas kelistrikan Universitas Brescia menunjukkan bahwa model gabungan ini unggul dalam menangkap pola kompleks dan meningkatkan akurasi prediksi. Poin penting dari penelitian ini adalah integrasi variabel eksternal, khususnya suhu udara, yang terbukti meningkatkan performa model secara signifikan. Temuan ini menunjukkan kemampuan *Random Forest Regression* untuk mengakomodasi variabel eksternal yang penting, sehingga meningkatkan fleksibilitas model dalam berbagai kondisi lingkungan.

Studi lainnya dilakukan oleh Marzak, dkk. (2024) untuk mengevaluasi pengaruh perilaku konsumen dan perubahan musim terhadap hasil *forecasting* permintaan beban listrik jangka pendek dengan memasukkan variabel historis dari kondisi cuaca dan data kalender. Mereka melakukan optimalisasi model melalui teknik *feature engineering* dan tuning *hyperparameter* dan menghasilkan tingkat akurasi *forecasting* yang tinggi berdasarkan metrik *Mean Absolute Percentage Error* (MAPE). Hal ini mengindikasikan bahwa model mampu mengidentifikasi interaksi dinamis antara fitur-fitur yang digunakan dan mampu menangani kompleksitas data yang dipengaruhi oleh perilaku manusia dan perubahan iklim musiman.

Sementara itu, Liu (2024) melakukan studi terkait *Random Forest Regression* dengan menggunakan berbagai fitur input seperti temperatur, kelembapan, kecepatan angin, dan hari libur. Selain itu mereka juga menerapkan teknik *resampling* yaitu validasi silang dan *grid search cross validation*. Studi mereka menghasilkan model *forecasting* yang akurat dan dapat diandalkan sebagai pendukung keputusan bagi perusahaan listrik, terutama dalam menghadapi lonjakan beban dan pengaturan *resource allocation* yang lebih efisien. Hal ini semakin menekankan kontribusi *Random Forest Regression* dalam pengelolaan fluktuasi permintaan listrik serta peningkatan efisiensi sistem tenaga secara keseluruhan.

Selanjutnya, Zhu dan Wu (2021) menegaskan fleksibilitas *Random Forest Regression* dengan membuktikan bahwa model ini lebih unggul dibanding metode *machine learning* lain dalam hal koefisien determinasi (R^2), serta memiliki nilai *Mean Absolute Error* (MAE) dan *Mean Squarred Error* (MSE) yang rendah. Selain itu, studi ini menunjukkan bahwa dengan menambahkan fitur dari sistem *Heating, Ventilation, and Air Conditioning* (HVAC), akurasi *forecasting* dapat ditingkatkan. Hal ini menunjukkan bahwa *Random Forest Regression* tidak hanya fleksibel dalam struktur modelnya, tetapi juga adaptif terhadap berbagai jenis variabel eksternal yang memengaruhi beban listrik, seperti penggunaan pendingin udara pada musim panas.

Studi oleh Johannesen, dkk. (2019) juga menegaskan bahwa *Random Forest Regression* memiliki keunggulan dalam menangani sejumlah besar fitur yang relevan untuk prediksi beban listrik, seperti kondisi cuaca, variabel kalender, dan data historis. Melalui pendekatan vertikal *time axis*, studi ini membandingkan performa *Random Forest Regression* dengan *k-Nearest Neighbour* (KNN) dan *Linear Regression*, dan menunjukkan bahwa *Random Forest Regression* unggul dalam prediksi jangka pendek (30 menit) dengan nilai *Mean Absolute Percentage Error* (MAPE) yang lebih rendah. Salah satu nilai tambah utama dari *Random Forest Regression* yang diidentifikasi dalam penelitian ini adalah kemampuannya dalam menghasilkan *feature importance*, yang memungkinkan pengguna untuk mengetahui kontribusi relatif masing-masing fitur terhadap hasil prediksi. Hal ini menjadikan *Random Forest Regression* lebih mudah untuk diinterpretasikan dibandingkan model lain yang tidak memiliki mekanisme interpretasi bawaan.

Hasil kajian dari berbagai studi tersebut menegaskan secara konsisten bahwa *Random Forest Regression* memiliki kemampuan dalam menangani kompleksitas data, fleksibilitas terhadap berbagai kondisi input, serta kontribusinya terhadap efisiensi dan keandalan sistem energi listrik. Hal ini menjadikan *Random Forest Regression* sebagai salah satu model paling ideal untuk digunakan dalam *forecasting* beban listrik jangka pendek.

Namun demikian, meskipun *Random Forest Regression* memiliki tingkat akurasi yang lebih baik dibanding model klasik dan cenderung lebih mudah diinterpretasikan daripada model *black-box*, interpretasi yang diberikan masih bersifat global dan deskriptif. Mekanisme *built-in feature importance* dalam *Random Forest Regression* yang umumnya dihitung berdasarkan rata-rata penurunan impurity atau kontribusi terhadap pengurangan error, tidak menunjukkan secara spesifik bagaimana fitur memengaruhi hasil prediksi pada level individual. Hal ini menjadikan interpretasi model *Random Forest Regression* kurang intuitif bagi pengambil keputusan non-teknis, serta menyulitkan dalam menjawab pertanyaan *what-if* atau menyusun skenario intervensi yang relevan secara operasional.

2.1.2 Interpretasi Hasil *Forecasting* Beban Listrik

Dalam beberapa tahun terakhir, upaya untuk meningkatkan interpretabilitas model prediktif melalui pendekatan *eXplainable AI* (XAI) telah mengalami perkembangan signifikan, termasuk dalam *forecasting* beban listrik. *Random Forest Regression* yang memiliki *built-in feature importance* mampu menghasilkan interpretasi yang menjelaskan kontribusi relatif dari masing-masing fitur terhadap hasil *forecasting*. Namun, beberapa studi menunjukkan bahwa pendekatan ini masih memiliki keterbatasan cukup signifikan, terutama ketika menggunakan data yang memiliki kompleksitas tinggi seperti data beban listrik.

Studi oleh Chandna (2023) menunjukkan bahwa *built-in feature importance* dapat terpengaruh oleh fenomena *curse of dimensionality*, di mana peningkatan jumlah fitur justru menyebabkan penurunan kemampuan generalisasi model. Dalam *forecasting* beban listrik, hal ini menjadi sangat relevan karena data historis yang digunakan biasanya bersifat multivariat dan kompleks. Ketika terlalu banyak fitur dimasukkan tanpa pendekatan yang hati-hati, model cenderung *overfitting* terhadap data pelatihan dan kehilangan kemampuan untuk memberikan *forecasting* yang akurat pada data baru. Selain itu, Chandna juga menggarisbawahi bahwa proses seleksi fitur berdasarkan skor *importance* dapat menyebabkan hilangnya informasi penting. Hal ini dikarenakan fitur-fitur dalam sistem energi sering kali saling berinteraksi secara kompleks, sehingga menyingkirkan fitur tertentu dapat mengabaikan efek sinergis yang tidak ditangkap oleh evaluasi individu fitur secara terpisah.

Sementara itu, Wallace, dkk. (2023) mengkritisi *built-in feature importance* dari sudut pandang metodologis yang lebih tajam, dengan menyoroti adanya *bias terhadap fitur yang saling berkorelasi*. Dalam praktiknya, fitur yang memiliki korelasi tinggi satu sama lain cenderung memperoleh skor penting yang tinggi, padahal kontribusi unik masing-masing fitur terhadap prediksi sebenarnya tidak sebesar yang ditampilkan. Hal ini dapat mengarahkan peneliti atau pengambil keputusan pada kesimpulan yang salah mengenai faktor-faktor utama dalam prediksi. Wallace dkk. mendemonstrasikan hal ini dalam studi prediksi risiko stroke, di mana dua variabel fungsi paru-paru yang berkorelasi tinggi muncul

sebagai fitur paling penting, menutupi peran signifikan dari faktor risiko medis yang sebenarnya lebih relevan secara klinis. Mereka menyarankan penggunaan pendekatan modern seperti *knockoff VIMPs*, yang secara statistik lebih valid dan mampu membedakan pengaruh kausal dari fitur terhadap model. Studi ini menekankan pentingnya kehati-hatian dalam menginterpretasi skor pentingnya fitur, terutama dalam pengambilan keputusan yang berdampak luas.

Di sisi lain, Prabhu, dkk. (2023) menyoroti permasalahan dari perspektif interpretabilitas praktis. Mereka mengakui bahwa meskipun *Random Forest* mampu menghasilkan skor pentingnya fitur, namun interpretasi terhadap skor tersebut bersifat terbatas dan tidak menjawab pertanyaan mendasar "*mengapa model memprediksi demikian untuk suatu instance tertentu?*". Dalam studi mereka yang berfokus pada prediksi ukuran tetesan dalam alat pemrosesan industri, penulis menyatakan bahwa hasil *feature importance* tidak cukup untuk memberikan informasi yang dibutuhkan untuk pengambilan keputusan teknis. Oleh karena itu, mereka mengintegrasikan pendekatan Local Interpretable.

Di sisi lain, teknik interpretasi model-agnostik seperti SHAP dan LIME telah banyak diterapkan dalam *forecasting* untuk menjelaskan kontribusi fitur terhadap hasil prediksi. Namun, teknik-teknik ini umumnya bersifat deskriptif dan belum mampu memberikan informasi yang bersifat preskriptif atau *actionable*. Mereka tidak menunjukkan skenario perubahan nilai input yang dapat mengubah output model ke arah yang diinginkan, serta cenderung kurang efisien pada data berdimensi tinggi atau ketika terjadi interaksi non-linier yang kompleks antar fitur.

Pendekatan *Counterfactual Explanations* menawarkan alternatif yang lebih preskriptif. *Counterfactual Explanations* menghasilkan skenario 'what-if' yang menunjukkan perubahan minimal pada fitur input agar hasil prediksi berubah ke arah yang diharapkan. Studi oleh Z. Wang, dkk. (2023) memperkenalkan pendekatan *ForecastCF*, sebuah teknik *Counterfactual Explanations* berbasis gradien untuk model *deep forecasting*, yang menunjukkan bahwa modifikasi kecil pada input historis dapat menghasilkan perubahan signifikan pada output prediksi. Sementara itu, Zuin dan Veloso (2024) mengembangkan pendekatan *Counterfactual Explanations* berbasis algoritma genetika yang mempertimbangkan

Granger causality dalam data *multivariate time series*, serta menerapkan *quantile regression* untuk menjaga realisme skenario yang dihasilkan.

Meskipun kedua pendekatan tersebut menunjukkan hasil yang menjanjikan, penerapannya dalam *forecasting* beban listrik masih terbatas. Selain itu, tantangan dalam menjamin keterkaitan kausal antar fitur masih menjadi hambatan terutama pada *ForecastCF* yang ditawarkan oleh Z. Wang, dkk karena pada studi mereka data yang digunakan merupakan data *time series* berjenis univariat.

2.1.3 Framework DiCE untuk *Counterfactual Explanations*

Dalam pengembangan model prediktif, interpretabilitas menjadi aspek krusial terutama untuk menjembatani pemahaman antara sistem berbasis machine learning dan pemangku kepentingan non-teknis. Pendekatan berbasis *feature importance* lebih sering digunakan, meskipun lebih bersifat statis dan tidak selalu memberikan konteks aksi yang jelas. Sebaliknya, *Counterfactual Explanations* menawarkan perspektif yang lebih interaktif dengan menyajikan skenario "what-if", dimana hal tersebut dapat menjelaskan bagaimana perubahan pada input tertentu dapat menghasilkan prediksi yang berbeda. Pendekatan ini memungkinkan pengguna mengevaluasi dampak perubahan fitur secara langsung dan memahami hubungan kausal dalam model, menjadikannya alat interpretasi yang lebih aplikatif dan informatif dalam mendukung pengambilan keputusan.

You, dkk. (2023) menunjukkan bahwa *Counterfactual Explanations* dapat memperkaya interpretasi model prediktif dengan memberikan informasi kausal yang sering kali tidak dapat ditangkap oleh skor *feature importance*. Mereka menekankan bahwa *Counterfactual Explanations* mampu menelusuri sebab-akibat antara fitur dan target dengan menghasilkan data sintetis yang secara minimal memodifikasi fitur untuk membalikkan hasil prediksi. Pendekatan ini menjadikan model lebih transparan dan dapat ditelaah di sekitar *decision boundaries*-nya.

Selaras dengan itu, Piccialli, dkk. (2024) memanfaatkan *Counterfactual Explanations* sebagai alat untuk mendeteksi batas keputusan dari model "*black-box*" dan kemudian menyusunnya kembali menjadi *decision tree* yang lebih mudah dipahami. Pendekatan ini menunjukkan bahwa *Counterfactual Explanations* tidak

hanya sekadar memberikan informasi secara lokal, tetapi juga dapat dimanfaatkan untuk rekonstruksi model secara global dalam bentuk yang lebih *interpretable*. Implikasi dari pendekatan ini sangat relevan untuk sistem prediksi beban listrik, di mana pemahaman terhadap batas keputusan sangat penting untuk mitigasi risiko dan pengelolaan sumber daya energi secara efisien.

Studi awal yang dilakukan Mothilal, dkk. (2020) memperdalam manfaat penggunaan *Counterfactual Explanations* dalam menginterpretasi sebuah model prediksi. Studi mereka memperkenalkan konsep penjelasan yang beragam yaitu *Diverse Counterfactual Explanations* (DiCE). Menurut mereka, keberagaman ini bukan hanya fitur tambahan, melainkan syarat penting agar pengguna mendapatkan berbagai opsi tindakan. Dengan menyuguhkan lebih dari satu solusi, keberagaman ini dapat meningkatkan kepercayaan terhadap sistem AI dan menghindarkan pengguna dari asumsi bahwa hanya ada satu cara untuk mencapai perubahan hasil. Penelitian mereka juga menunjukkan bahwa pendekatan ini mampu mengoptimalkan keseimbangan antara *realisme*, *kelayakan tindakan (actionability)*, dan *diversitas*. Kinerja ini tidak hanya efektif secara teknis, tetapi juga siap diterapkan dalam konteks dunia nyata yang kompleks dan menuntut solusi cepat.

Keberagaman *Counterfactual Explanations* ini juga dinilai penting dalam lingkup etika dan keadilan. Slack, dkk. (2021) menyoroti risiko manipulasi yang dapat terjadi jika sistem hanya memberikan satu *Counterfactual Explanations*. Dalam eksperimen yang dilakukan, mereka menemukan bahwa ketergantungan pada satu penjelasan dapat membuka celah eksploitasi, terutama pada sistem-sistem yang berdampak langsung pada individu seperti pinjaman atau prediksi kejahatan. Dengan menyuguhkan banyak skenario alternatif maka akan dapat meningkatkan ketangguhan (*robustness*) sistem terhadap manipulasi dan membuat penjelasan lebih adil antar subkelompok data.

Lebih lanjut, Mothilal, dkk. (2021) melakukan studi empiris dengan membandingkan kualitas interpretasi yang dihasilkan oleh teknik XAI berbasis *feature importance* dengan teknik *Counterfactual Explanations*. Mereka menyatakan bahwa meskipun teknik berbasis *feature importance* seperti SHAP atau

LIME dapat mengidentifikasi fitur penting, mereka gagal menjawab pertanyaan *"bagaimana saya bisa mengubah hasil prediksi?"* dengan cara yang konkrit dan dapat ditindaklanjuti. Kommiya menunjukkan bahwa pendekatan *Counterfactual Explanations*, khususnya yang bersifat diverse seperti DiCE, dapat melengkapi bahkan melampaui *Feature Importance* dalam hal memberikan skenario 'what-if' yang masuk akal dan kontekstual. Mereka berpendapat bahwa *Counterfactual Explanations* dan *Feature Importance* sebenarnya mengarah pada tujuan yang sama—yakni menjelaskan perilaku model—namun dari pendekatan dan manfaat yang berbeda. *Counterfactual Explanations* lebih fokus pada perubahan minimal dan implikatif terhadap input, sehingga lebih cocok digunakan dalam proses pengambilan keputusan operasional dan strategi kebijakan.

Berdasarkan temuan dari berbagai studi tersebut, framework DiCE menjadi pilihan yang strategis dalam mengembangkan sistem *forecasting* beban listrik yang tidak hanya akurat, tetapi juga dapat dijelaskan, dipahami, dan digunakan secara aktif oleh pengambil keputusan. Integrasi DiCE memungkinkan pengguna untuk menavigasi hasil prediksi secara eksploratif, tidak hanya secara pasif, dengan memahami berbagai jalan menuju perubahan dan mengambil tindakan berdasarkan skenario yang paling sesuai dengan konteks dan kebijakan yang berlaku.

2.2 Dasar Teori

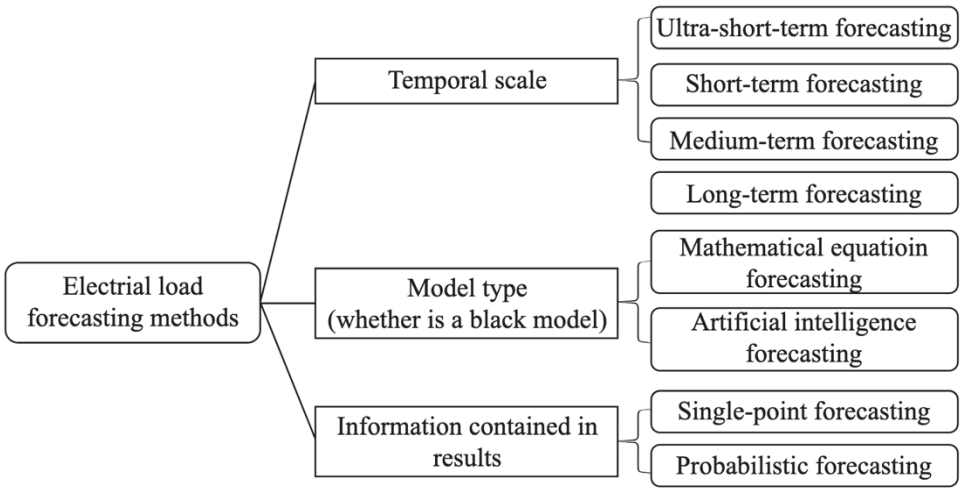
2.2.1 *Forecasting* Beban Listrik

Transformasi sistem tenaga listrik global menuju arah yang bersih, rendah karbon, dan cerdas mendorong perlunya pendekatan peramalan beban listrik yang akurat dan adaptif. Perubahan struktural dalam pembangkitan dan konsumsi energi, seperti peningkatan integrasi sumber energi terbarukan dan penerapan sistem respons permintaan, memperbesar ketidakpastian pada sisi suplai dan permintaan. Untuk menghadapi kondisi ini, sistem tenaga listrik membutuhkan metode peramalan yang tidak hanya mampu mengantisipasi dinamika beban, tetapi juga dapat mendukung pengambilan keputusan operasional yang cepat dan tepat.

Forecasting beban listrik menjadi komponen penting dalam manajemen sistem tenaga karena secara langsung memengaruhi kegiatan perencanaan,

pengoperasian, dan pengendalian jaringan. Ketepatan prediksi beban dapat meningkatkan efisiensi pembangkitan, mengurangi biaya cadangan daya, serta meningkatkan keandalan distribusi energi. Kesalahan dalam *forecasting* dapat berdampak sistemik, seperti pembebanan jaringan yang tidak seimbang, pemborosan energi, atau bahkan pemadaman. Oleh karena itu, teknik *forecasting* harus mampu menyesuaikan diri dengan kompleksitas dan ketidakpastian dalam sistem tenaga yang dinamis.

Menurut H. Wang, dkk. (2022), metode *forecasting* beban listrik dapat dikelompokkan berdasarkan tiga kriteria yang saling melengkapi, yaitu skala temporal prediksi (*temporal scale*), jenis model yang digunakan (*model type*), dan bentuk informasi yang dihasilkan dari hasil prediksi (*information contained in results*). Ketiga kriteria ini tidak berdiri secara terpisah, melainkan membentuk dimensi klasifikasi yang dapat digunakan secara simultan dalam mendeskripsikan karakteristik suatu metode *forecasting*. Struktur klasifikasi ini dapat dilihat pada Gambar 2.1.



Gambar 2.1 Pengelompokan Metode *Forecasting* Beban Listrik.

Berdasarkan skala waktu (*temporal scale)*, *forecasting* beban listrik dibagi menjadi empat kelompok, yaitu *ultra-short-term*, *short-term*, *medium-term*, dan *long-term forecasting*. Masing-masing kategori mencerminkan cakupan waktu prediksi, yang pada gilirannya menentukan tujuan dan strategi operasional sistem.

Misalnya, peramalan jangka sangat pendek (dalam hitungan menit hingga jam) umumnya digunakan untuk kontrol real-time, sedangkan peramalan jangka panjang (tahunan) digunakan dalam perencanaan infrastruktur dan kebijakan energi.

Dari sisi pendekatan pemodelan (*model type)*, metode *forecasting* dibagi menjadi *mathematical equation-based models* yang meliputi pendekatan statistik dan persamaan deterministik dan *artificial intelligence-based models* yang mencakup berbagai algoritma pembelajaran mesin dan jaringan saraf. Pemilihan model sangat bergantung pada kompleksitas sistem, ketersediaan data historis, serta kebutuhan interpretabilitas dan akurasi.

Berdasarkan bentuk informasi yang dihasilkan (*information contained in result)*, metode *forecasting* dibagi menjadi *single-point forecasting* dan *probabilistic forecasting*. Pendekatan *single-point* memberikan satu nilai estimasi spesifik, sedangkan pendekatan *probabilistic* menghasilkan distribusi atau rentang prediksi dengan tingkat keyakinan tertentu, yang sangat berguna dalam pengelolaan risiko dan ketidakpastian sistem tenaga.

Analisis dan pengembangan metode *forecasting* beban listrik dapat dilakukan secara lebih sistematis dan kontekstual jika merujuk pada ketiga kriteria ini. Pemilihan kombinasi yang tepat dari ketiga kriteria ini akan sangat menentukan efektivitas model dalam mendukung pengambilan keputusan operasional maupun strategis dalam sistem tenaga listrik modern yang kompleks dan dinamis.

Selain pemilihan metode, pemilihan dan pengolahan variabel input juga sangat mempengaruhi tingkat akurasi *forecasting* beban listrik. Variabel eksternal seperti suhu, kelembapan, hari libur, dan sebagainya memiliki hubungan signifikan terhadap fluktuasi beban. Oleh karena itu, integrasi informasi cuaca, kalender operasional, serta pola konsumsi historis menjadi elemen penting dalam membangun model prediksi yang andal dan responsif terhadap perubahan (H. Wang dkk., 2022).

2.2.2 *Random Forest Regression*

Random Forest Regression merupakan salah satu algoritma *machine learning* berbasis *ensemble learning* yang bekerja dengan menggabungkan banyak

model prediktif sederhana berupa pohon keputusan (*decision tree*). Tujuannya adalah menghasilkan prediksi yang lebih akurat dan stabil. Metode ini merupakan pengembangan dari *Decision Tree Regression* dengan menerapkan pendekatan *bagging* serta seleksi acak fitur, guna mengatasi kelemahan utama dari pohon keputusan tunggal seperti overfitting dan sensitivitas terhadap fluktuasi data pelatihan.

Dalam penerapannya, *Random Forest Regression* membangun sekumpulan pohon keputusan untuk melakukan prediksi pada tugas regresi. Setiap pohon dibangun menggunakan subset data dan subset fitur yang dipilih secara acak, sehingga masing-masing pohon memiliki struktur dan hasil prediksi yang berbeda. Prediksi akhir dari model diperoleh dengan cara mengagregasi hasil prediksi dari seluruh pohon, umumnya melalui rata-rata. Pendekatan ini bertujuan meningkatkan akurasi sekaligus mengurangi variansi, dua komponen penting dalam keseimbangan *bias-variance trade-off*.

Proses pembentukan model dimulai dengan teknik *bootstrap sampling*, yaitu mengambil sampel acak dari dataset pelatihan sebanyak N data dengan pengembalian. Hal ini memungkinkan sebagian data muncul lebih dari sekali dalam satu subset, dan sebagian lainnya tidak muncul sama sekali. Proses ini diulang untuk setiap pohon yang akan dibangun, menghasilkan variasi dataset pelatihan antar pohon dalam forest.

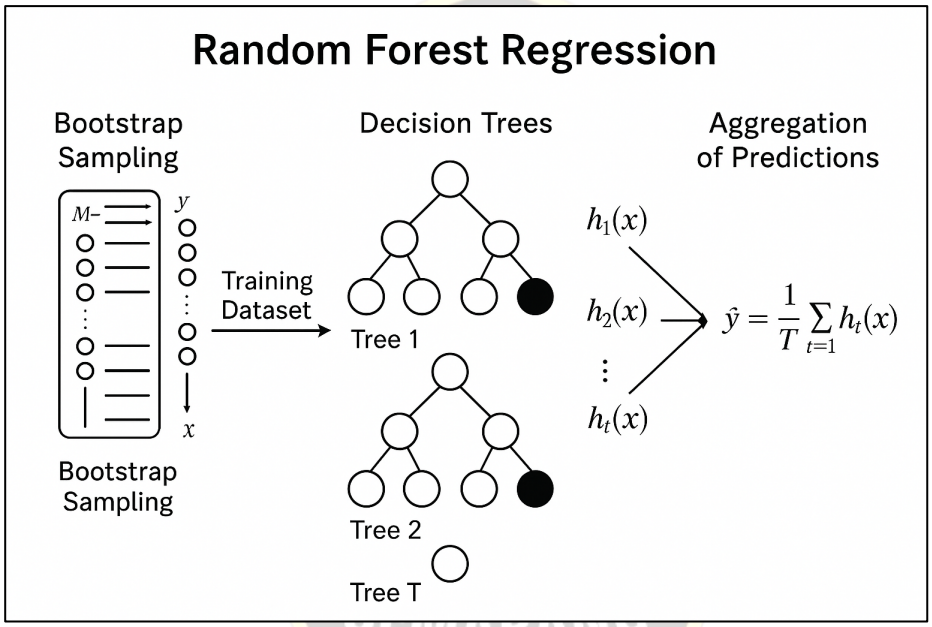
Pada saat membentuk simpul dalam pohon keputusan, hanya sebagian fitur yang dipilih secara acak dari seluruh fitur yang tersedia. Jika total fitur adalah M, maka hanya m fitur dengan $m<M$ yang dipertimbangkan untuk setiap proses pemisahan. Seleksi acak ini meningkatkan keberagaman antar pohon dan membantu mengurangi korelasi antar pohon dalam ensemble.

Setiap pohon kemudian dibangun sedalam mungkin tanpa proses pemangkasan (*pruning*), hingga mencapai simpul akhir atau *leaf node*. Karena Random Forest terdiri dari banyak pohon, tidak diperlukan proses pengendalian kompleksitas pada masing-masing pohon secara eksplisit. Sebaliknya, overfitting dicegah secara keseluruhan melalui proses agregasi.

Setelah seluruh pohon selesai dibangun, prediksi untuk data uji x' dihitung dengan mengambil rata-rata hasil dari semua pohon. Secara matematis, prediksi akhir dituliskan sebagai:

$$\hat{y}(x') = \frac{1}{T}\sum_{t=1}^{T} h_1(x') \tag{2.1}$$

di mana $h_1(x')$ adalah prediksi dari pohon ke-*t*, dan *T* adalah jumlah total pohon dalam forest. Ilustrasi proses kerja *Random Forest Regression* dapat dilihat pada Gambar 2.2.



Gambar 2.2 Ilustrasi Proses Kerja *Random Forest Regression*.

Berdasarkan Gambar 2.2 proses kerja *Random Forest Regression* dimulai dari data pelatihan yang dibagi secara acak menggunakan teknik *bootstrap sampling*, lalu masing-masing pohon dibangun menggunakan subset data dan subset fitur acak. Hasil prediksi dari semua pohon digabungkan (rata-rata) untuk menghasilkan prediksi akhir.

Random Forest Regression memiliki sejumlah keunggulan penting. Salah satunya adalah tingkat stabilitas yang tinggi. Dengan menggabungkan banyak pohon yang dibangun dari kombinasi data dan fitur yang berbeda, model menjadi

lebih tahan terhadap perubahan kecil dalam data pelatihan. Selain itu, algoritma ini mampu memberikan performa yang baik pada dataset berdimensi tinggi serta mampu menangani hubungan *non-linear* antara fitur dan target. Hal ini sangat relevan dalam konteks seperti *forecasting* beban listrik, di mana pola data yang rumit dan dinamis menjadi tantangan tersendiri.

Keunggulan lain yang menonjol adalah kemampuannya dalam menilai pentingnya fitur (*feature importance*). Model dapat memberikan estimasi kontribusi relatif dari masing-masing fitur terhadap performa prediksi secara keseluruhan, yang sangat bermanfaat dalam proses interpretasi hasil. Beberapa implementasi Random Forest juga mampu menangani nilai hilang dengan baik, baik pada saat pelatihan maupun saat inferensi.

Dalam *forecasting* beban listrik, *Random Forest Regression* memiliki keunggulan dalam menangani dataset yang besar, kompleks, dan bersifat non-linear. Algoritma ini mampu menangkap interaksi variabel yang kompleks tanpa asumsi distribusi data tertentu. *Random Forest Regression* telah memberikan hasil yang kompetitif dan stabil dibandingkan dengan metode peramalan tradisional maupun beberapa model machine learning lainnya, terutama dalam horizon peramalan jangka pendek hingga menengah (Lahouar dan Slama, 2015).

2.2.3 *eXplainable AI* dan *Counterfactual Explanations*

*Explainable AI (*XAI*)* merupakan bidang dalam kecerdasan buatan (Artificial Intelligence/AI) yang berfokus pada pengembangan sistem yang mampu memberikan penjelasan yang dapat dimengerti manusia atas keputusan atau prediksi yang dihasilkan oleh model. Kebutuhan akan XAI muncul sebagai respons terhadap sifat black-box dari banyak model AI modern, seperti Random Forest, XGBoost, dan Deep Neural Networks, yang meskipun menawarkan tingkat akurasi tinggi, sering kali sulit dipahami bagaimana dan mengapa model menghasilkan suatu keluaran tertentu.

Dalam berbagai domain kritikal seperti kesehatan, keuangan, hukum, dan sistem energi, ketidakmampuan untuk menjelaskan keputusan model dapat menghambat kepercayaan pengguna, mengurangi transparansi, serta mengaburkan

akuntabilitas sistem secara keseluruhan. Oleh karena itu, kehadiran XAI menjadi penting sebagai jembatan antara kinerja prediktif model dan kebutuhan pemahaman manusia. Menurut Barredo Arrieta, dkk. (2020) sistem XAI yang ideal adalah sistem yang tidak hanya memberikan penjelasan teknis, tetapi juga menjelaskan prediksi dalam bentuk yang dapat diterima secara sosial dan psikologis oleh pengguna akhir.

Molnar (2019) membedakan metode interpretabilitas menjadi dua kategori. Yang pertama adalah model yang secara intrinsik dapat dijelaskan atau *interpretable models*, yaitu model yang dirancang sejak awal agar logika internalnya transparan dan mudah diikuti oleh manusia, seperti decision tree dan regresi linear. Kategori kedua adalah penjelasan pasca-pelatihan atau *post-hoc explanations* yang dapat diterapkan setelah model selesai dilatih. Pendekatan ini bertujuan menjelaskan perilaku model yang kompleks tanpa mengubah struktur internalnya, melalui visualisasi, estimasi kontribusi fitur, atau simulasi perilaku model terhadap input yang dimodifikasi. Salah satu metode post-hoc yang berkembang pesat adalah *Counterfactual Explanations*.

Kualitas penjelasan yang diberikan oleh sistem XAI sangat bergantung pada siapa pengguna akhirnya. Menurut Miller (2019) penjelasan yang baik bukan hanya menjelaskan hubungan sebab-akibat secara objektif, melainkan juga harus relevan secara kognitif bagi pengguna. Penjelasan yang efektif cenderung bersifat kontrasif, yaitu membandingkan antara hasil aktual dengan hasil yang tidak terjadi namun mungkin; bersifat selektif, yakni hanya menyoroti faktor-faktor paling berpengaruh; serta bersifat sosial, dalam artian disampaikan dalam konteks komunikasi yang sesuai dengan tingkat pemahaman audiens.

Dalam kerangka inilah *Counterfactual Explanations* menjadi pendekatan yang intuitif dan natural. *Counterfactual Explanations* bertujuan menjelaskan prediksi model dengan menunjukkan bagaimana perubahan kecil dan minimal pada input dapat mengubah hasil prediksi. Pendekatan ini merefleksikan cara berpikir manusia dalam memahami sebab-akibat, yaitu dengan mempertanyakan apa yang akan terjadi jika kondisi awal berbeda. Misalnya, apabila model memprediksi bahwa beban listrik pada hari esok berada pada tingkat tinggi, *Counterfactual*

Explanations dapat memberikan informasi berupa perubahan seperti pengurangan suhu maksimum sebesar dua derajat Celsius dan penurunan konsumsi dasar sebesar lima persen, yang akan menyebabkan prediksi beban menjadi lebih rendah. Penjelasan seperti ini tidak hanya menggambarkan alasan di balik keputusan model, tetapi juga menyarankan intervensi nyata yang dapat dilakukan untuk mencapai hasil yang diinginkan, sehingga bersifat actionable.

Counterfactual Explanations memiliki tiga karakteristik. Pertama, bersifat kontrasif karena menjelaskan mengapa suatu hasil terjadi dibandingkan dengan hasil alternatif yang mungkin. Kedua, bersifat sosial karena bertujuan menjembatani komunikasi antara sistem dan pengguna, sehingga penjelasan harus disesuaikan dengan kebutuhan dan pemahaman audiens. Ketiga, bersifat selektif karena hanya menyoroti sejumlah faktor penting yang relevan dengan perubahan hasil.

Walaupun *Counterfactual Explanations* telah banyak digunakan dalam konteks klasifikasi, penerapannya dalam konteks *forecasting*, terutama peramalan deret waktu (*time series forecasting*), masih tergolong baru dan menghadirkan tantangan tersendiri. Peramalan deret waktu umumnya menghasilkan serangkaian prediksi pada horizon waktu tertentu, bukan hanya satu nilai keluaran. Selain itu, struktur data bersifat sekuensial, di mana satu titik waktu dipengaruhi oleh dinamika titik waktu sebelumnya. Kompleksitas ini menuntut pendekatan *Counterfactual Explanations* yang lebih cermat agar tetap relevan dan valid secara temporal.

Z. Wang dkk. (2023) menyatakan bahwa tantangan utama dalam *Counterfactual Explanations* untuk *forecasting* terletak pada bagaimana memodifikasi urutan historis dari deret waktu agar prediksi masa depan sesuai dengan target yang diinginkan. Untuk itu, mereka menetapkan sejumlah prinsip yang harus dipenuhi oleh *Counterfactual Explanations* yang baik. Pertama, hasil prediksi dari *Counterfactual Explanations* harus valid secara fungsional, artinya nilai prediksi yang baru harus berada dalam batas yang diharapkan, seperti tidak melampaui ambang batas beban atau mengikuti tren penurunan yang dirancang. Kedua, perubahan terhadap data historis harus dekat dengan *manifold* data asli,

yakni tidak menyimpang dari pola aktual seperti musiman atau tren jangka panjang, sehingga hasil prediksi tetap masuk akal secara domain. Ketiga, perubahan yang dilakukan harus minimal, baik dari sisi jumlah titik waktu yang dimodifikasi maupun dari besarnya nilai perubahan, sehingga tetap efisien dan mudah diimplementasikan. Keempat, *Counterfactual Explanations* harus dapat ditindaklanjuti dan dapat diinterpretasikan, dalam artian bahwa perubahan nilai input dapat dimaknai sebagai tindakan nyata seperti penyesuaian operasi, pengaturan ulang jadwal, atau modifikasi faktor lingkungan yang relevan.

Wang dan timnya juga menunjukkan visualisasi perubahan beberapa titik historis dalam deret waktu, misalnya hari ke-10 hingga ke-14 dari total 14 hari historis, yang menghasilkan prediksi baru pada tujuh hari ke depan yang lebih sesuai dengan target operasional. Untuk menilai kualitas *Counterfactual Explanations* yang dihasilkan, mereka mengusulkan beberapa metrik evaluasi, seperti rasio validitas (*validity ratio*) yang menunjukkan proporsi prediksi masa depan yang sesuai target, area di bawah kurva validitas bertahap (*stepwise validity AUC*) untuk mengukur konsistensi prediksi, ukuran kedekatan (*proximity*) antara input asli dan *Counterfactual Explanations*, serta ukuran kesederhanaan (*compactness*) berdasarkan proporsi titik waktu yang tidak diubah.

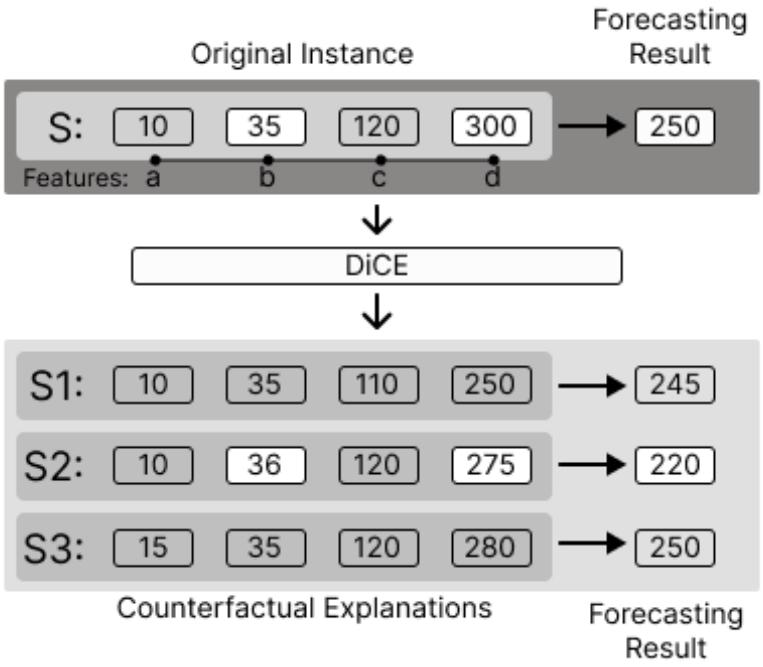
Melalui pendekatan ini, *Counterfactual Explanations* dapat memberikan penjelasan yang kontrasif, menyeluruh, dan dapat ditindaklanjuti dalam konteks *forecasting*. Penjelasan ini tidak hanya menjawab pertanyaan "mengapa suatu prediksi muncul", tetapi juga "apa yang dapat diubah untuk mengubah hasil", yang sangat penting dalam sistem yang memerlukan pengambilan keputusan cepat dan tepat, seperti dalam pengelolaan beban listrik.

2.2.4 *Diverse Counterfactual Explanations* (DiCE) *Framework*

Salah satu tantangan utama dalam penerapan *Counterfactual Explanations* adalah bagaimana menghasilkan contoh *Counterfactual Explanations* yang tidak hanya valid dan realistis, tetapi juga dapat dilakukan secara efisien pada model *black-box* yang kompleks. Untuk menjawab kebutuhan ini, sebuah *framework*

bernama *Diverse Counterfactual Explanations* (DiCE) yang berbasis optimasi gradien untuk menemukan *Counterfactual Explanations* dikembangkan.

DiCE adalah framework yang dirancang untuk menghasilkan kumpulan penjelasan *Counterfactual Explanations* yang valid, beragam, dan efisien secara komputasional. DiCE dikembangkan dengan mengacu pada prinsip bahwa satu keputusan model dapat dijelaskan melalui berbagai jalur kemungkinan yang sah. Pendekatan ini sangat berbeda dengan metode *Counterfactual Explanations* sebelumnya yang hanya menghasilkan satu solusi optimal, karena kenyataannya dalam pengambilan keputusan, pengguna seringkali membutuhkan beberapa alternatif yang dapat dibandingkan dan dipilih sesuai dengan kendala atau kebijakan yang berlaku. Ilustrasi proses pembentukan *Counterfactual Explanations* menggunakan DiCE ditampilkan pada Gambar 2.3.



Gambar 2.3 Ilustrasi Proses Kerja Pembentukan *Counterfactual Explanations* Menggunakan DiCE.

Secara teknis, DiCE menggunakan metode optimisasi multi-objektif untuk menghasilkan kumpulan *Counterfactual Explanations* yang memaksimalkan validitas prediksi, kedekatan terhadap data asli (*proximity*), dan keragaman antar penjelasan (*diversity*). Mothilal, dkk. (2020) menyusun fungsi objektif yang

menggabungkan metrik-metrik tersebut untuk menghasilkan *Counterfactual Explanations* yang tidak hanya berbeda dari prediksi awal, tetapi juga berbeda satu sama lain, sehingga pengguna memperoleh wawasan yang lebih luas tentang jalur intervensi yang tersedia. Selain itu, DiCE dapat diintegrasikan dengan batasan domain (*domain constraints*) untuk memastikan bahwa hasil *Counterfactual Explanations* tetap realistis, misalnya: "suhu tidak boleh negatif", atau "waktu tidak boleh melompat mundur".

Keunggulan utama dari DiCE terletak pada fleksibilitas dan sifat *model-agnostic* yang dimilikinya. DiCE dapat diterapkan pada berbagai jenis model *machine learning*, termasuk model *black-box* seperti *Random Forest Regression*, tanpa perlu mengetahui struktur internal model. Pendekatan ini juga mendukung integrasi dengan pustaka populer seperti *scikit-learn*, *TensorFlow*, dan *PyTorch* melalui implementasi Python dalam pustaka dice-ml. Dalam evaluasi kuantitatif yang dilakukan oleh Mothilal, dkk. (2021), DiCE menunjukkan keunggulan dalam menghasilkan *Counterfactual Explanations* yang lebih valid dan beragam dibandingkan pendekatan berbasis *feature importance*.

Dalam *forecasting* beban listrik, DiCE memberikan nilai tambah interpretatif yang signifikan. Dengan menerapkan DiCE, pengguna dapat memperoleh berbagai skenario "*what-if*" terkait faktor-faktor seperti suhu, hari operasional, atau konsumsi historis, yang berdampak pada hasil *forecasting*. Misalnya, sistem dapat menghasilkan alternatif seperti: "Jika suhu maksimum kemarin 3°C lebih rendah *atau* konsumsi pada dua hari terakhir dikurangi 10%, maka beban puncak esok hari akan berada dalam batas aman." Penjelasan semacam ini tidak hanya menjelaskan hasil model, tetapi juga menyediakan pilihan tindakan nyata yang dapat dijadikan dasar pengambilan keputusan oleh operator sistem tenaga listrik.

Melalui pendekatan yang adaptif dan *user-centric*, DiCE dapat menjembatani kesenjangan antara kekuatan prediktif model AI dan kebutuhan interpretabilitas pengguna akhir. Framework ini memposisikan penjelasan sebagai alat eksplorasi dan pengambilan keputusan, bukan sekadar pelengkap hasil *forecasting*. Oleh karena itu, DiCE menjadi salah satu solusi interpretabilitas yang

sangat relevan dalam implementasi sistem *forecasting* beban listrik yang transparan, bertanggung jawab, dan dapat ditindaklanjuti.

2.2.5 Metrik Evaluasi *Time Series Forecasting*

Dalam melakukan evaluasi kinerja sebuah model, digunakan beberapa metrik yang dapat mengukur seberapa dekat nilai *forecasting* yang dihasilkan oleh sebuah model terhadap nilai asli.

Metrik *Mean Absolute Error* (MAE), *Root Mean Squared Error* (RMSE), dan *Mean Absolute Percentage Error* (MAPE) adalah metrik yang biasa digunakan dalam evaluasi model *forecasting*. Masing-masing metrik tersebut memiliki karakteristik dan tujuan tersendiri untuk menilai kinerja dan akurasi sebuah model *forecasting*.

2.2.5.1 MAE

MAE adalah metrik evaluasi yang mengukur rata-rata dari selisih absolut antara nilai aktual dan nilai hasil *forecast*. Metrik ini bersifat linier dimana setiap kesalahan akan memiliki bobot yang sama tanpa memperhitungkan arah kesalahan tersebut baik positif maupun negatif.

Metrik MAE dinilai cukup intuitif dan sering digunakan sebagai evaluasi dasar karena dapat memberikan informasi secara langsung mengenai rata-rata kesalahan yang terjadi.

Untuk menghitung nilai MAE, digunakan persamaan sebagai berikut:

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|y_t - \hat{y}_t| \quad (2.2)$$

dimana n merupakan jumlah total observasi dalam dataset, y_t adalah nilai aktual atau asli dari target waktu ke-t dan $\hat{y}_t$ adalah nilai hasil forecast pada waktu ke-t. Tanda mutlak digunakan untuk memastikan semua nilai ditetapkan dalam rentang nilai positif.

2.2.5.2 RMSE

RMSE menghitung akar kuadrat dari rata-rata selisih kuadrat antara nilai aktual dan *forecast*. Tidak seperti MAE, metrik ini memberikan penalti lebih besar terhadap kesalahan yang lebih besar karena menggunakan pendekatan kuadrat. Oleh karena itu, RMSE sangat berguna ketika evaluasi model difokuskan untuk menghindari kesalahan prediksi besar.

Perhitungan nilai RMSE menggunakan persamaan berikut:

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^{n} (y_t - \hat{y}_t)^2} \tag{2.3}$$

dimana dalam persamaan diatas, y_t dan $\hat{y}_t$ tetap merepresentasikan nilai aktual dan nilai prediksi pada waktu ke-t, dan n adalah jumlah observasi. Selisih antara nilai aktual dan prediksi dikuadratkan terlebih dahulu untuk menghindari nilai negatif, lalu dirata-rata dan diakar kuadratkan untuk mengembalikannya ke satuan yang sama dengan data asli. Hal ini membuat RMSE tetap sensitif terhadap outlier dan lebih mudah diinterpretasikan.

2.2.5.3 MAPE

Metrik MAPE dianggap metrik yang cukup intuitif dan mudah dipahami oleh pengguna non-teknis, karena menyajikan tingkat kesalahan dari model dalam bentuk persentase.

Metrik MAPE menggunakan persamaan berikut untuk menghasilkan nilai persentase dari tingkat kesalahan model:

$$MAPE = \frac{100\%}{n} \sum_{t=1}^{n} \left| \frac{y_t - \hat{y}_t}{y_t} \right| \tag{2.4}$$

dimana selisih antara nilai aktual dan nilai prediksi dibagi dengan nilai aktual pada waktu ke-t, kemudian diambil nilai absolutnya agar tidak saling meniadakan. Nilai tersebut kemudian dirata-ratakan dan dikalikan dengan 100 untuk mendapatkan hasil dalam satuan persen.

Penggunaan metrik MAPE perlu dicermati secara hati-hati terutama jika terdapat nilai aktual yang sangat kecil atau bahkan nol, karena akan menghasilkan nilai MAPE yang sangat besar atau bahkan tidak terdefinisi dan menyebabkan kesalahan interpretasi hasil evaluasi.

2.2.6 Metrik Evaluasi *Counterfactual Explanations*

2.2.6.1 Validity

Metrik ini mengukur keberhasilan skenario *Counterfactual Explanations* dalam menghasilkan perubahan pada output model sesuai dengan target yang diinginkan. Dalam sistem *forecasting* beban listrik, tujuan interpretasi adalah memahami bagaimana menurunkan prediksi beban listrik di bawah ambang batas tertentu. Sehingga *Counterfactual Explanations* akan dianggap valid jika prediksi hasil dari skenario tersebut memang menunjukkan penurunan di bawah ambang tersebut.

Validity dihitung sebagai proporsi dari jumlah contoh counterfactual unik yang dihasilkan yang benar-benar memberikan prediksi kelas yang berbeda (kelas yang diinginkan) dibandingkan dengan input asli (Mothilal dkk., 2020). Sehingga metrik *validity* menggunakan persamaan berikut untuk mendapatkan nilai proporsi jumlah *Counterfactual Explanations* yang berhasil mengubah prediksi sesuai dengan target yang ditentukan:

$$Validity = \frac{1}{\mathrm{m}} \sum_{\mathrm{i}=1}^{\mathrm{m}} \mathbb{I}(f(x_i') \geq t) \tag{2.5}$$

dimana untuk setiap hasil prediksi yang dihasilkan dalam skenario *Counterfactual Explanations* yaitu $f(x_i')$ diverifikasi apakah telah memenuhi ambang batas target t. Setiap keberhasilan diberi bobot 1 melalui fungsi indikator, kemudian dirata-ratakan terhadap seluruh jumlah counterfactual yang dihasilkan.

Metrik ini memastikan bahwa *Counterfactual Explanations* yang dihasilkan tidak hanya merupakan variasi input tanpa makna, melainkan benar-benar mampu mempengaruhi output model secara signifikan dan sesuai tujuan.

2.2.6.2 Proximity

Metrik ini mengukur jarak *euclidean* rata-rata antara nilai instance asli dan instance *Counterfactual Explanations* (Z. Wang dkk., 2023). Metrik ini penting untuk memastikan bahwa perubahan yang disarankan tetap realistis dan berada dalam ruang kemungkinan yang wajar secara domain. Oleh karena itu, *proximity* yang rendah (perubahan minimal) menunjukkan bahwa skenario yang dihasilkan tidak hanya efektif, tetapi juga praktis untuk dipertimbangkan dalam pengambilan keputusan.

Metrik *proximity* menggunakan persamaan berikut untuk mendapatkan jarak nilai *Counterfactual Explanations* terhadap nilai asli:

$$Proximity_i = \sqrt{\sum_{j=1}^{n} \left(x_j^{orig} - x_j^{CF_i}\right)^2} \tag{2.6}$$

$$\overline{Proximity} = \frac{1}{m} \sum_{i=1}^{m} Proximity_i \tag{2.7}$$

2.2.6.3 Compactness

Dalam sistem *forecasting* beban listrik, interpretasi yang terlalu kompleks atau melibatkan banyak perubahan variabel sekaligus cenderung sulit diterjemahkan ke dalam kebijakan atau tindakan operasional. Melalui metrik ini, pengukuran terhadap seberapa banyak fitur yang perlu diubah untuk menghasilkan perubahan output model dapat dilakukan.

Counterfactual Explanations yang baik, idealnya hanya memerlukan sedikit perubahan fitur agar hasil model berubah secara signifikan. Hal ini membuat interpretasi menjadi lebih mudah dipahami oleh pengguna dan lebih sederhana untuk diimplementasikan dalam praktik(Mothilal dkk., 2020).

Untuk mendapatkan nilai dari metrik *compactness*, digunakan persamaan berikut:

$$\text{Compactness} = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(x_i' \neq x_i) \tag{2.8}$$

dimana jumlah fitur yang berubah atau berbeda dari nilai asli $\mathbb{I}(x_i' \neq x_i)$, dibagi dengan total fitur yang tersedia n.

Dalam beberapa literatur, metrik ini sering dikenal juga dengan nama *sparsity* yang mengukur jumlah fitur yang berbeda antara nilai instance asli dan nilai skenario *Counterfactual Explanations* yang dihasilkan. Nilai yang lebih rendah menunjukkan bahwa lebih sedikit fitur yang perlu dimodifikasi, sehingga menghasilkan penjelasan yang lebih sederhana dan lebih mudah diinterpretasikan.