

BAB I

PENDAHULUAN

1.1. Latar Belakang

Perkembangan teknologi informasi telah meningkatkan kemampuan organisasi dalam menyimpan, mengelola, dan menganalisis data. Data yang sebelumnya hanya berfungsi sebagai catatan administratif kini dapat diolah menjadi informasi yang mendukung pengambilan keputusan. Salah satu pendekatan yang dapat digunakan dalam proses tersebut adalah *machine learning*. Pendekatan ini memungkinkan komputer mempelajari pola hubungan antarvariabel berdasarkan data dan menggunakan pola tersebut untuk menghasilkan prediksi terhadap konsumsi rata rata pelanggan di kabupaten labuhanbatu untuk dapat menurunkan resiko kelebihan penggunaan listrik yang tidak terpakai.

Pemanfaatan analisis data memiliki peran penting dalam pengelolaan konsumsi energi listrik. Setiap pelanggan memiliki karakteristik yang berbeda, seperti merek meter kWh, golongan tarif, kapasitas daya, kecamatan, dan kelurahan. Perbedaan karakteristik tersebut dapat berkaitan dengan variasi konsumsi listrik rata-rata antarpelanggan. Data pelanggan yang hanya disajikan dalam bentuk tabel belum secara langsung menjelaskan pola hubungan antara karakteristik pelanggan dan tingkat konsumsi listriknya. Oleh karena itu, data tersebut perlu dianalisis menggunakan metode yang mampu mengenali pola hubungan antarvariabel secara sistematis.

Pelanggan listrik di Kabupaten Labuhanbatu tersebar pada sejumlah kecamatan dan kelurahan dengan karakteristik yang beragam. Pelanggan yang menggunakan kapasitas daya sama belum tentu memiliki konsumsi listrik rata-rata yang sama karena dapat memiliki golongan tarif, merek meter kWh, dan lokasi yang berbeda. Sebaliknya, pelanggan yang berada pada wilayah administratif yang sama juga dapat menunjukkan tingkat konsumsi berbeda karena mempunyai karakteristik pelanggan yang tidak sama. Kondisi tersebut menunjukkan bahwa konsumsi listrik rata-rata tidak cukup dianalisis hanya berdasarkan satu karakteristik.

Data yang digunakan dalam penelitian ini terdiri atas 6.510 observasi pelanggan. Variabel prediktor meliputi merek meter kWh, golongan tarif, daya dalam satuan volt-ampere, kecamatan, dan kelurahan. Variabel target yang diprediksi adalah konsumsi listrik rata-rata setiap pelanggan. Setiap baris data merepresentasikan karakteristik dan nilai konsumsi listrik rata-rata satu pelanggan dalam satu kumpulan data.

Data penelitian ini bersifat nonhistoris karena tidak disusun berdasarkan urutan waktu bulanan atau tahunan pada setiap pelanggan. Data tidak memuat rangkaian pengamatan konsumsi dari waktu ke waktu yang dapat digunakan untuk menganalisis tren atau pola musiman. Dengan demikian, prediksi dalam penelitian ini tidak diartikan sebagai peramalan konsumsi listrik pada bulan atau tahun mendatang. Prediksi diartikan sebagai proses memperkirakan nilai konsumsi listrik rata-rata pelanggan berdasarkan pola hubungan antara karakteristik pelanggan dan nilai konsumsi yang dipelajari oleh model.

Karakteristik data tersebut menjadi dasar dalam menentukan ruang lingkup dan batas interpretasi hasil penelitian. Penelitian ini tidak ditujukan untuk memperkirakan pertumbuhan kebutuhan energi listrik, perubahan konsumsi tahunan, atau beban listrik pada masa mendatang. Penelitian difokuskan pada variasi antarpelanggan dalam satu kumpulan data. Oleh karena itu, hasil model hanya dapat digunakan untuk menjelaskan kemampuan karakteristik pelanggan yang tersedia dalam memperkirakan konsumsi listrik rata-rata.

Masalah utama penelitian ini adalah belum diketahui seberapa baik variabel merek meter kWh, golongan tarif, daya, kecamatan, dan kelurahan dapat digunakan untuk memperkirakan konsumsi listrik rata-rata pelanggan. Variabel daya berbentuk numerik, sedangkan merek meter kWh, golongan tarif, kecamatan, dan kelurahan berbentuk kategorikal. Variasi jenis data tersebut dapat membentuk pola hubungan yang kompleks. Hubungan antara karakteristik pelanggan dan konsumsi listrik rata-rata juga tidak harus berbentuk linear. Interaksi beberapa karakteristik dapat menghasilkan pola konsumsi berbeda pada kelompok pelanggan tertentu.

Kondisi tersebut memerlukan metode regresi yang mampu mengenali pola nonlinier dan interaksi antarvariabel tanpa menetapkan bentuk hubungan matematis tertentu sejak awal. Salah satu algoritma yang dapat digunakan adalah *Random Forest Regression*. Algoritma ini membangun sejumlah pohon regresi dari variasi sampel data dan variabel yang dipilih secara acak. Setiap pohon menghasilkan nilai prediksi, kemudian hasil seluruh pohon dirata-ratakan untuk

memperoleh prediksi akhir. Penggabungan banyak pohon bertujuan menghasilkan model yang lebih stabil dibandingkan penggunaan satu pohon keputusan.

Penerapan *Random Forest Regression* dalam penelitian ini dilakukan melalui tahapan pengolahan data yang sistematis. Data terlebih dahulu diperiksa untuk mengetahui kesesuaian tipe variabel, kelengkapan nilai, serta kemungkinan adanya data yang tidak dapat digunakan. Variabel kategorikal kemudian diubah ke dalam bentuk numerik agar dapat diproses oleh model. Setelah tahap prapemrosesan, data dibagi menjadi data latih dan data uji. Data latih digunakan untuk membangun model, sedangkan data uji digunakan untuk menilai kemampuan model dalam memprediksi observasi yang tidak dilibatkan dalam proses pelatihan.

Kinerja model dievaluasi menggunakan *Mean Absolute Error* (MAE), *Mean Squared Error* (MSE), *Root Mean Squared Error* (RMSE), koefisien determinasi atau (R^2), dan *Symmetric Mean Absolute Percentage Error* (sMAPE). MAE menunjukkan rata-rata kesalahan absolut antara nilai aktual dan nilai prediksi. MSE menghitung rata-rata kesalahan yang telah dikuadratkan sehingga memberikan penalti lebih besar terhadap kesalahan bernilai besar. RMSE menunjukkan besarnya kesalahan prediksi dalam satuan yang sama dengan variabel target. Nilai (R^2) menunjukkan kemampuan model dalam menjelaskan variasi konsumsi listrik pada data uji. Sementara itu, sMAPE menunjukkan besar kesalahan prediksi secara relatif dan simetris.

Hasil evaluasi *Random Forest Regression* perlu dibandingkan dengan model dasar atau *baseline*. Model *baseline* menghasilkan prediksi menggunakan

pendekatan sederhana tanpa mempelajari pola kompleks dari variabel prediktor. Perbandingan tersebut diperlukan untuk mengetahui apakah penerapan Random Forest memberikan kemampuan prediktif yang lebih baik. Model dapat dinilai memberikan manfaat apabila menghasilkan nilai kesalahan yang lebih rendah atau kemampuan menjelaskan variasi yang lebih baik daripada *baseline* pada data uji yang sama.

Penelitian ini juga menggunakan *permutation importance* untuk mengukur kontribusi relatif setiap variabel terhadap kinerja model. Pengukuran dilakukan dengan mengacak nilai satu variabel pada data evaluasi, kemudian mengamati perubahan kinerja prediksi. Penurunan kinerja yang lebih besar menunjukkan bahwa variabel tersebut memiliki peran lebih penting dalam proses prediksi model. Hasil analisis ini dapat memberikan informasi mengenai karakteristik pelanggan yang paling informatif dalam memperkirakan konsumsi listrik rata-rata. Namun, tingkat kepentingan variabel hanya menunjukkan kontribusi terhadap model dan tidak membuktikan hubungan sebab-akibat.

Penerapan *Random Forest* pada data pelanggan di Kabupaten Labuhanbatu diharapkan menghasilkan gambaran empiris mengenai kemampuan karakteristik pelanggan dalam memperkirakan konsumsi listrik rata-rata. Nilai prediksi menunjukkan hasil estimasi model terhadap konsumsi pelanggan. Nilai metrik evaluasi menunjukkan tingkat kesalahan dan kemampuan model dalam menjelaskan variasi data. Sementara itu, hasil *permutation importance* menunjukkan kontribusi relatif masing-masing variabel prediktor terhadap kinerja model.

Nilai khusus penelitian ini terletak pada penerapan *Random Forest Regression* terhadap data pelanggan yang bersifat nonhistoris. Penelitian tidak menggunakan urutan waktu sebagai dasar prediksi, tetapi memanfaatkan variasi karakteristik antarpelanggan. Fokus penelitian terletak pada kemampuan algoritma dalam memetakan hubungan antara merek meter kWh, golongan tarif, daya, kecamatan, dan kelurahan dengan konsumsi listrik rata-rata pelanggan. Dengan demikian, hasil penelitian harus ditafsirkan sebagai estimasi berbasis karakteristik pelanggan, bukan sebagai peramalan konsumsi listrik pada masa mendatang.

Penelitian dilakukan untuk membangun model *Random Forest Regression* yang dapat memperkirakan konsumsi listrik rata-rata setiap pelanggan berdasarkan karakteristik yang tersedia. Penelitian ini juga bertujuan mengukur kinerja model menggunakan MAE, MSE, RMSE, (R^2), dan sMAPE, membandingkan kinerja *Random Forest* dengan model *baseline*, serta mengidentifikasi kontribusi relatif setiap variabel prediktor menggunakan *permutation importance*. Seluruh hasil penelitian dibatasi pada karakteristik data nonhistoris dan tidak digunakan untuk menyimpulkan perubahan konsumsi listrik berdasarkan waktu.

1.2. Rumusan Masalah

1. Bagaimana penerapan algoritma *Random Forest* dalam membangun model prediksi konsumsi listrik pelanggan di Kabupaten Labuhanbatu berdasarkan karakteristik pelanggan yang tersedia?

2. Bagaimana kinerja model *Random Forest Regression* dalam memprediksi konsumsi listrik rata-rata pelanggan prabayar berdasarkan *Mean Absolute Error* (MAE), *Root Mean Squared Error* (RMSE), *Mean Squared Error* (MSE), koefisien determinasi (R^2), dan *Symmetric Mean Absolute Percentage Error* (sMAPE) pada data uji?
3. Apakah model *Random Forest* mampu menghasilkan prediksi konsumsi listrik yang lebih baik dibandingkan model *baseline* berdasarkan hasil pengujian pada data uji?

1.3. Ruang Lingkup Masalah

1. Penelitian menggunakan 6.510 data pelanggan listrik prabayar yang berasal dari enam kecamatan di Kabupaten Labuhanbatu, yaitu Rantau Selatan, Rantau Utara, Bilah Hulu, Bilah Barat, Pangkatan, dan Bilah Hilir.
2. Variabel prediktor dibatasi pada merek meter kWh, golongan tarif, daya terpasang, kecamatan, dan kelurahan. Variabel target berupa nilai konsumsi listrik rata-rata pelanggan. Kolom ID_RECORD hanya digunakan sebagai identitas data dan tidak dimasukkan sebagai variabel prediktor.
3. Penelitian menggunakan *Random Forest Regression*. Kinerja model dinilai dengan MAE, MSE, RMSE, R^2 , dan sMAPE serta dibandingkan dengan model *baseline*. Prediksi dibatasi pada estimasi konsumsi listrik rata-rata berdasarkan karakteristik pelanggan, bukan peramalan konsumsi berdasarkan bulan atau tahun.

1.4. Tujuan dan Manfaat Penelitian

1.4.1. Tujuan Penelitian

1. Menerapkan algoritma *Random Forest Regression* untuk membangun model prediksi konsumsi listrik rata-rata pelanggan di Kabupaten Labuhanbatu berdasarkan merek meter kWh, golongan tarif, daya terpasang, kecamatan, dan kelurahan.
2. Mengetahui kinerja model *Random Forest Regression* dalam memprediksi konsumsi listrik rata-rata pelanggan berdasarkan nilai *Mean Absolute Error* (MAE), *Mean Squared Error* (MSE), *Root Mean Squared Error* (RMSE), koefisien determinasi (R^2), dan *Symmetric Mean Absolute Percentage Error* (sMAPE).
3. Membandingkan kinerja model *Random Forest Regression* dengan model baseline untuk mengetahui apakah *Random Forest* mampu memberikan hasil prediksi yang lebih baik pada data uji.

1.4.2. Manfaat Penelitian

1. Penelitian ini diharapkan menambah bukti empiris mengenai penerapan *Random Forest Regression* untuk memprediksi konsumsi listrik rata-rata berdasarkan karakteristik pelanggan pada data *nonhistoris*. Penelitian ini juga memberikan gambaran tentang kemampuan dan keterbatasan *Random Forest* pada data yang bersifat cross-sectional serta memperjelas perbedaan antara prediksi lintas pelanggan dan peramalan runtut waktu.

2. Hasil penelitian diharapkan memberikan gambaran awal mengenai pemanfaatan data merek meter kWh, golongan tarif, daya terpasang, kecamatan, dan kelurahan dalam memperkirakan konsumsi listrik rata-rata pelanggan. Hasil tersebut dapat menjadi bahan pertimbangan awal dalam pengembangan analisis pelanggan berbasis data.
3. Penelitian ini diharapkan menjadi referensi untuk mengembangkan model prediksi konsumsi listrik dengan variabel yang lebih lengkap, riwayat konsumsi berkala, cakupan wilayah yang lebih representatif, serta algoritma pembandingan. Temuan dan keterbatasan penelitian juga dapat membantu peneliti selanjutnya menyusun model yang lebih kuat.

1.5. Tinjauan Umum Objek Penelitian

Objek penelitian adalah data pelanggan listrik yang teridentifikasi berada di Kabupaten Labuhanbatu, Provinsi Sumatera Utara. Kabupaten Labuhanbatu memiliki sembilan kecamatan secara administratif, tetapi data lengkap yang tersedia untuk penelitian ini mencakup enam kecamatan. Batas tersebut harus dinyatakan secara jelas agar hasil penelitian tidak dianggap mewakili pelanggan yang tidak tercakup dalam dataset.

Unit analisis penelitian adalah pelanggan listrik. Setiap pelanggan mempunyai satu nilai target berupa konsumsi rata-rata listrik dan sejumlah karakteristik prediktor. Struktur ini berbeda dari data deret waktu. Model tidak mempelajari urutan bulan, tren tahunan, atau pola musiman. Model mempelajari hubungan antara variasi karakteristik pelanggan dan variasi nilai konsumsi pada kumpulan data yang tersedia.

Data lengkap terdiri atas 6.510 pelanggan, dua golongan tarif, dua kelompok daya, tujuh merek kWh meter, enam kecamatan, dan 45 kelurahan. Variasi kategori tersebut memberikan informasi bagi model untuk membedakan karakteristik pelanggan. Namun, dominasi kategori tertentu tetap perlu diperhatikan karena dapat membuat model lebih akurat pada kelompok mayoritas daripada kelompok yang jumlahnya kecil.

Penelitian menjaga kerahasiaan pelanggan dengan mengeluarkan identitas langsung. Data yang digunakan untuk pemodelan hanya memuat atribut yang relevan. Pendekatan ini sejalan dengan prinsip minimalisasi data, yaitu menggunakan hanya informasi yang diperlukan untuk menjawab tujuan penelitian.

1.6. Sistematika Penulisan

Sistematika penulisan penelitian ini disusun secara terstruktur agar setiap tahapan penelitian dapat dipahami secara logis dan sistematis. Penyusunan setiap bab diselaraskan dengan rumusan masalah, ruang lingkup, tujuan penelitian, metode yang digunakan, serta karakteristik data penelitian. Laporan penelitian ini terdiri atas lima bab dengan sistematika sebagai berikut.

BAB I PENDAHULUAN. Bab ini menjelaskan dasar pelaksanaan penelitian. Pembahasannya mencakup latar belakang, identifikasi masalah, rumusan masalah, ruang lingkup masalah, tujuan penelitian, manfaat penelitian, tinjauan umum objek penelitian, dan sistematika penulisan. Bab ini memberikan gambaran mengenai fokus penelitian, yaitu penerapan algoritma Random Forest untuk memprediksi konsumsi listrik rata-rata pelanggan di Kabupaten Labuhanbatu.

BAB II LANDASAN TEORI. Bab ini menguraikan teori dan konsep yang mendukung penelitian. Pembahasan meliputi teori dan konsep data, prediksi, *machine learning*, *supervised learning*, *regresi*, *decision tree*, *Random Forest Regression*, pengolahan data, pengodean variabel kategorikal, pembagian data latih dan data uji, penyetelan hyperparameter, serta evaluasi model menggunakan MAE, MSE, RMSE, R^2 , dan sMAPE. Bab ini juga memuat penelitian terdahulu dan kerangka berpikir sebagai dasar ilmiah penelitian.

BAB III METODE PENELITIAN. Bab ini menjelaskan tahapan penelitian secara terperinci. Pembahasannya mencakup lokasi dan waktu penelitian, jenis penelitian, sumber data, objek penelitian, variabel penelitian, tahapan pengolahan data, analisis eksploratif, pembagian data, pengodean variabel kategorikal, pembangunan model baseline, penerapan Random Forest Regression, penyetelan *hyperparameter*, dan pengujian model. Bab ini juga menjelaskan prosedur yang digunakan untuk mencegah kebocoran data dan menjaga objektivitas evaluasi.

BAB IV HASIL DAN PEMBAHASAN. Bab ini menyajikan hasil pengolahan data dan penerapan *Random Forest Regression*. Pembahasan mencakup karakteristik dataset, hasil praproses, konfigurasi model, hasil pelatihan dan pengujian, serta nilai MAE, MSE, RMSE, R^2 , dan sMAPE. Kinerja *Random Forest* dibandingkan dengan model baseline untuk mengetahui apakah model memberikan hasil prediksi yang lebih baik. Bab ini juga membahas hasil penelitian berdasarkan teori, penelitian terdahulu, dan keterbatasan data.

BAB V KESIMPULAN DAN SARAN. Bab ini memuat kesimpulan dan saran. Kesimpulan disusun berdasarkan hasil penelitian untuk menjawab setiap rumusan masalah. Saran diberikan kepada pihak terkait dan peneliti selanjutnya dengan mempertimbangkan hasil pengujian, cakupan data, keterbatasan variabel, serta peluang pengembangan model prediksi konsumsi listrik.