Random forest result FI

Berdasarkan data yang anda tunjukkan, terdapat dua tahap di mana anda menggunakan Feature Importance: satu sebelum modelling (semasa feature selection) dan satu lagi selepas modelling (selepas melatih model).

**Feature Importance Sebelum Modelling (Seleksi Ciri)**

Di sini, anda menggunakan Feature Importance untuk memilih ciri yang paling penting berdasarkan nilai yang lebih tinggi. Anda memilih ciri-ciri yang mempunyai nilai Importance lebih tinggi daripada purata (mean), dan menggunakan ciri-ciri tersebut untuk membina model. Berikut adalah ciri-ciri yang anda pilih:

* **Contract** (0.200376)
* **tenure** (0.168621)
* **MonthlyCharges** (0.116200)
* **OnlineSecurity** (0.098771)
* **TotalCharges** (0.092103)
* **TechSupport** (0.074966)

Ini menunjukkan bahawa sebelum modelling, ciri-ciri seperti **Contract**, **tenure**, dan **MonthlyCharges** adalah antara yang paling penting dalam menjelaskan model anda. Pemilihan ciri ini boleh membantu untuk mengurangkan dimensi data, seterusnya meningkatkan prestasi model.

**Feature Importance Selepas Modelling (Model Performance)**

Selepas anda membina dan melatih model, anda kembali menganalisis ciri-ciri penting yang dihasilkan oleh model tersebut. Hasil Feature Importance selepas modelling memberi gambaran yang lebih terperinci mengenai peranan sebenar setiap ciri dalam menjelaskan ramalan model.

**Perbandingan Antara Kedua-Dua:**

* **Ciri-ciri Yang Kekal Penting:**
  + **MonthlyCharges** (0.239185) dan **TotalCharges** (0.226259) menjadi dua ciri paling penting selepas modelling, berbanding dengan seleksi awal.
  + **tenure** (0.196061) juga tetap menjadi ciri penting selepas modelling, walaupun sedikit turun daripada kedudukan awal.
* **Perubahan Ciri yang Ditambah:**
  + **Contract\_Month-to-month** dan **Contract\_Two year** (nilai lebih tinggi selepas one-hot encoding) muncul sebagai ciri penting dalam model yang dilatih.
  + Ciri-ciri seperti **TechSupport\_No** dan **OnlineSecurity\_No** juga mula muncul sebagai lebih penting berbanding semasa seleksi ciri.

**Perubahan dalam Encoding:**

1. **Label Encoding vs One-Hot Encoding:**
   * **Label Encoding** adalah proses menukarkan kategori kepada nombor (contoh: "Yes" = 1, "No" = 0). Ini lebih sesuai untuk pemboleh ubah yang mempunyai urutan tertentu (ordinal).
   * **One-Hot Encoding** menukarkan kategori menjadi lajur berasingan untuk setiap kategori (contoh: "Contract" boleh jadi "Contract\_One year", "Contract\_Two year", dll.). Ini lebih sesuai untuk pemboleh ubah yang tidak mempunyai urutan tertentu (nominal).

Apabila anda menggunakan **One-Hot Encoding**, ciri-ciri yang asalnya satu pemboleh ubah (seperti **Contract**) terbahagi kepada beberapa pemboleh ubah baru, seperti **Contract\_One year**, **Contract\_Two year**, dan **Contract\_Month-to-month**. Inilah sebab mengapa selepas **One-Hot Encoding**, ciri-ciri tersebut muncul sebagai lebih penting dalam model.

**Cara Untuk Presentasi:**

1. **Pengenalan:**
   * Mulakan dengan menjelaskan mengapa anda memilih Feature Importance dan bagaimana ia membantu dalam pemilihan ciri untuk model.
   * Terangkan bahawa anda menggunakan dua pendekatan: pertama dengan **Label Encoding** dan kedua dengan **One-Hot Encoding** selepas modelling.
2. **Tunjukkan Keputusan Sebelum Modelling:**
   * Paparkan keputusan Feature Importance sebelum modelling dan terangkan ciri-ciri yang anda pilih (seperti **Contract**, **tenure**, dan **MonthlyCharges**).
   * Nyatakan bahawa ini adalah langkah penting dalam memilih ciri yang berpotensi mempunyai impak terbesar terhadap model.
3. **Tunjukkan Keputusan Selepas Modelling:**
   * Paparkan hasil selepas modelling menggunakan **One-Hot Encoding** dan bandingkan dengan keputusan sebelumnya.
   * Jelaskan bagaimana beberapa ciri, seperti **Contract\_Month-to-month** dan **Contract\_Two year**, muncul sebagai penting selepas **One-Hot Encoding**.
   * Terangkan perbezaan dalam nilai Importance antara ciri yang dipilih semasa seleksi ciri dan ciri-ciri yang sebenarnya penting dalam model.
4. **Analisis Perubahan:**
   * Sampaikan perbandingan antara kedua-dua hasil.
   * Highlight bagaimana **One-Hot Encoding** mempengaruhi keputusan **Feature Importance** dan kenapa beberapa ciri yang penting menjadi lebih ketara selepas encoding.
5. **Kesimpulan:**
   * Akhiri dengan kesimpulan bahawa **One-Hot Encoding** memberi gambaran yang lebih jelas tentang hubungan antara ciri-ciri kategori dan sasaran model.
   * Nyatakan bahawa selepas proses encoding dan modelling, model lebih memberi keutamaan kepada ciri-ciri yang lebih relevan dalam menjelaskan output (seperti **MonthlyCharges**, **TotalCharges**, dan **tenure**).

**Ringkasan untuk Pembentangan:**

* **Langkah 1:** Gunakan Feature Importance untuk memilih ciri-ciri terbaik sebelum modelling.
* **Langkah 2:** Gunakan One-Hot Encoding untuk pemboleh ubah kategori.
* **Langkah 3:** Latih model dan dapatkan Feature Importance selepas modelling.
* **Langkah 4:** Bandingkan keputusan dan terangkan perbezaan.

Ini akan memberikan gambaran yang jelas tentang bagaimana setiap langkah mempengaruhi hasil dan mengapa One-Hot Encoding menghasilkan perubahan dalam model.