

Nama: Lea Audrey Felicia Sukamto

NIM: 2702261914

II. Essay

1. **Jelaskan rencana pengujian Anda secara sistematis dan Sertakan contoh berdasarkan fitur dalam dataset obesitas, seperti bagaimana jika nilai pada Age, Weight, SMOKE, atau MTRANS dimanipulasi.**

Jawaban:

1. Pengujian Fungsionalitas Model

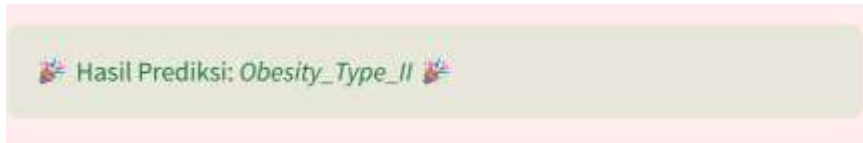
- **Tujuan:** Memastikan bahwa model dapat memproses data input secara benar dan memberikan output yang akurat.
- **Metode:** Uji dengan berbagai dataset untuk melihat apakah model menghasilkan prediksi yang konsisten dan akurat.
- **Contoh Pengujian:** Coba masukkan dataset dengan nilai usia (Age), berat badan (Weight), kebiasaan merokok (SMOKE), dan moda transportasi (MTRANS) yang berbeda. Misalnya:
 - **Age:** Coba dengan nilai ekstrem, seperti usia sangat muda (misalnya 5 tahun) atau usia sangat tua (misalnya 50 tahun).
 - **Weight:** Uji dengan berat badan yang sangat rendah (misalnya 30 kg) atau sangat tinggi (misalnya 150 kg).
 - **SMOKE:** Pastikan model dapat menangani kategori yang tidak terduga, seperti nilai "tidak merokok" atau "merokok".
 - **MTRANS:** Uji dengan kategori moda transportasi yang tidak lazim, seperti "berjalan kaki" atau "bersepeda" yang berbeda dari data pelatihan.

Contoh Pengujian:

- Age: 5, Weight: 30 kg, SMOKE: tidak, MTRans: berjalan kaki



- Age: 50, Weight: 150, SMOKE: ya, MTRans: Mobil



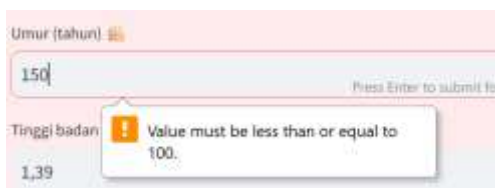
2. Pengujian Keamanan Data

- **Tujuan:** Memastikan bahwa model aman dari potensi penyalahgunaan atau manipulasi data input.
- **Metode:** Uji dengan data input yang terkontaminasi atau dimanipulasi, seperti nilai yang sangat tinggi atau sangat rendah yang dapat menyebabkan model berperilaku tidak semestinya.
- **Contoh Pengujian:** Coba masukkan nilai-nilai ekstrem atau invalid pada Age, Weight, atau MTRANS. Misalnya, nilai usia (Age) yang jauh di luar batas normal (misalnya 150 tahun), atau berat badan yang sangat tinggi (misalnya 1000 kg).

Contoh Pengujian:

Dalam Streamlit

- Age: 150



- Weight: 1000



Dimana user tidak bisa menginput nilai yang melebihi dari batas ketentuan, sehingga nantinya model tidak perlu memprediksi data yang anomaly.

3. Pengujian Pengaruh Manipulasi Data

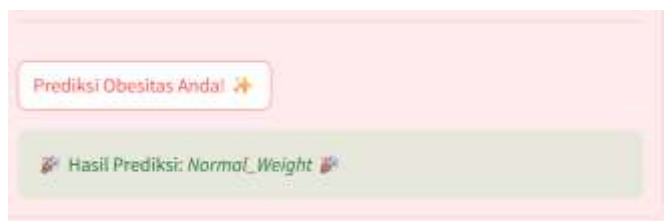
- **Tujuan:** Memastikan model tidak sensitif atau rentan terhadap perubahan minor atau manipulasi data input.
- **Metode:** Uji bagaimana model bereaksi terhadap perubahan kecil dalam nilai input, terutama pada fitur penting seperti Age, Weight, SMOKE, atau MTRANS.
- **Contoh Pengujian:** Coba manipulasi nilai pada fitur:
 - **Age:** Uji dengan sedikit perubahan pada usia, misalnya mengganti usia 35 menjadi 36 tahun, dan lihat apakah prediksi obesitas berubah drastis.
 - **Weight:** Coba ubah berat badan dari 70 kg menjadi 72 kg dan lihat apakah model merespons secara proporsional.
 - **SMOKE:** Coba ubah status merokok dari "merokok" menjadi "tidak merokok" dan lihat apakah prediksi tingkat obesitas terpengaruh.

Contoh Pengujian:

- Age: 35, Weight: 70, SMOKE: ya



- Age: 36, Weight: 75, SMOKE: tidak



Model merespon baik terhadap perubahan kecil.

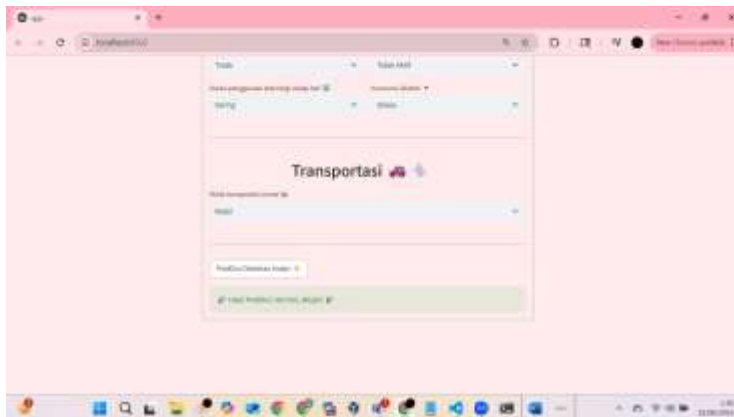
4. Pengujian Interoperabilitas

- **Tujuan:** Memastikan bahwa model dapat berjalan dengan lancar dalam berbagai sistem atau platform yang berbeda.
- **Metode:** Uji model dalam berbagai lingkungan produksi atau platform.

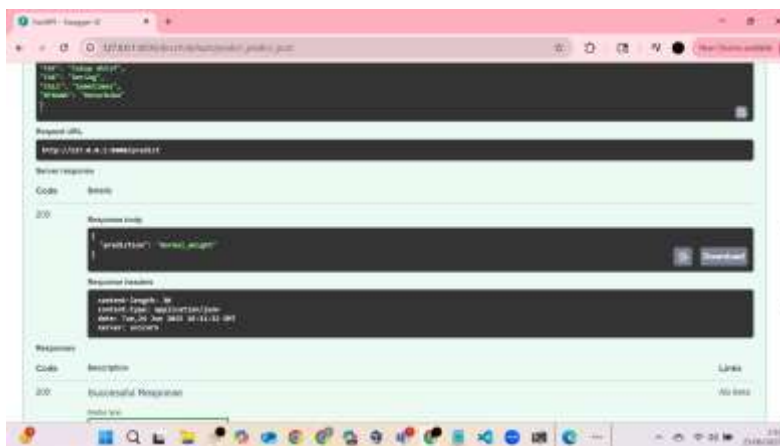
- **Contoh Pengujian:** Uji model di server yang berbeda atau dalam konteks aplikasi berbasis web untuk memastikan bahwa model dapat diintegrasikan dan berfungsi dengan benar tanpa gangguan.

Pengujian dalam 2 platform:

- Streamlit:



- FastAPI



2. Jelaskan apa yang dimaksud dengan model drift dan data drift dalam konteks sistem prediksi obesitas ini?

Jawaban:

- **Model Drift**

Model drift, atau yang juga dikenal dengan konsep *concept drift*, terjadi ketika ada perubahan dalam hubungan antara fitur input dan output prediksi seiring berjalannya waktu. Hal ini dapat terjadi jika distribusi atau pola data yang digunakan untuk melatih model berubah setelah model diimplementasikan di sistem produksi. Dengan kata lain, meskipun model bekerja dengan baik pada data pelatihan awal, ia bisa kehilangan efektivitasnya ketika konsep yang mendasari data berubah.

Dalam konteks prediksi obesitas, model drift bisa terjadi jika faktor-faktor seperti usia (Age), berat badan (Weight), kebiasaan merokok (SMOKE), atau moda transportasi (MTRANS) berhubungan secara berbeda dengan tingkat obesitas dari waktu ke waktu. Sebagai contoh, jika tren gaya hidup masyarakat berubah. Misalnya, semakin banyak orang yang mengadopsi pola makan sehat dan berolahraga maka model yang dilatih dengan data yang mencerminkan pola lama tidak lagi memberikan hasil yang akurat.

Contoh model drift dalam prediksi obesitas:

- Jika pada masa lalu, konsumsi makanan tinggi kalori berhubungan erat dengan obesitas, namun dalam beberapa tahun terakhir banyak orang yang mulai mengadopsi pola makan sehat, model yang dilatih menggunakan data lama mungkin tidak dapat lagi memprediksi tingkat obesitas secara tepat.

- Data Drift

Data drift terjadi ketika distribusi data yang diberikan kepada model berubah dari waktu ke waktu. Berbeda dengan model drift yang mengacu pada perubahan hubungan antara fitur dan output, data drift berfokus pada perubahan dalam data input itu sendiri. Meskipun model tetap sama, perubahan dalam karakteristik data yang diterima bisa menyebabkan penurunan akurasi prediksi.

Dalam sistem prediksi obesitas, data drift bisa terjadi jika ada perubahan dalam pola hidup masyarakat atau kebiasaan makan yang tercermin dalam data input, seperti usia (Age), berat badan (Weight), kebiasaan merokok (SMOKE), atau moda transportasi (MTRANS). Misalnya, jika pola makan cepat saji yang semula dominan mulai berkurang dan lebih banyak orang beralih ke pola makan sehat, maka distribusi data yang digunakan untuk

pelatihan model dapat berubah, sehingga model tidak lagi dapat memprediksi tingkat obesitas dengan akurat.

Contoh data drift dalam prediksi obesitas:

- Jika model dilatih dengan data dari kelompok masyarakat yang memiliki kebiasaan makan cepat saji, namun setelah beberapa waktu terjadi perubahan pola makan di masyarakat menuju pola yang lebih sehat, distribusi data seperti berat badan atau kebiasaan merokok yang digunakan oleh model bisa berubah. Hal ini dapat menyebabkan penurunan akurasi dalam memprediksi obesitas.

Sumber:

WebMD. (n.d.). *Health News*. Retrieved from <https://www.webmd.com/>

Rani, S., Yadav, A., & Soni, P. (2018). Concept Drift and Data Drift in Machine Learning. *International Journal of Computer Applications*, 181(10), 16-19. Retrieved from <https://www.ijcaonline.org/>

Google Cloud. (2020, December 3). *How to Handle Data Drift and Model Drift in Machine Learning*. Google Cloud Blog. Retrieved from <https://cloud.google.com/blog>