

Nama	: Adhimas Aryo Bimo
NIM	: 13523052

- Perbedaan jelas antara holdout validation dan Kfold pada pembagian data train dan testnya.
 - Hold-out validation
Data dibagi menjadi 2 sekaligus dan biasanya digunakan 80% untuk training dan 20% untuk testing. Hal ini cepat dan sederhana tapi hasil evaluasi bisa tergantung random state yang digunakan.
 - Kfold cross validation
Dataset dibagi menjadi k fold dengan ukuran yang relatif sama. Lalu, model dilatih sebanyak k kali dan hasil evaluasi pada semua fold dirata-ratakan diakhir. Hal ini sangat berguna untuk mengetahui bagaimana kualitas model pada setiap pembagian dataset
- Hold-out baik pada dataset yang sangat besar. Hal ini akan mempercepat dan anggapannya pada data besar, tiap pembagian untuk train dan test mendapatkan value yang representatif untuk setiap percobaan. Pada kfold kebalikannya, hal ini sangat berguna pada data yang kecil yang mungkin saja terdapat anomali pada data tersebut yang sulit untuk direpresentasikan jika pembagian train dan testnya menggunakan hold-out.

Kebalikannya Kfold buruk untuk dataset besar karena proses training yang akan sangat memakan waktu sedangkan hold-out akan buruk pada dataset kecil yang mungkin pembagiannya tidak representatif untuk semua kondisi.
- Singkatnya, ketika informasi dari data uji bocor ke data train. Misal, ketika melakukan scaling sebelum split antara train dan test. Atau variabel target yang tidak sengaja masuk ke fitur training.
- Dampaknya, model pada train akan sangat bagus, tetapi ketika diuji pada data sesungguhnya tidak memberikan hasil yang sesuai ketika train.
- Pastikan tidak ada variabel target pada training. Pastikan juga ketika scaling dilakukan setelah split data train dan test. Tidak hanya scaling, imputing juga termasuk.