

Classification and Regression Tree

1. Jelaskan cara kerja dari algoritma tersebut! (boleh dalam bentuk *pseudocode* ataupun narasi)

⇒ Pertama, model akan menerima data latihan. Dari data latihan tersebut, model akan membentuk sebuah pohon keputusan. Pohon keputusan tersebut terdiri dari beberapa *node*.

Node yang paling ujung merupakan hasil keputusan paling akhir untuk menentukan jenis klasifikasi. Sedangkan *node* yang lain berperan sebagai pengambil keputusan. Model akan melatih pengambilan keputusan tersebut dengan membuat berbagai cabang dan melihat apakah informasi yang didapatkan setelah pengambilan keputusan tersebut sudah bagus atau belum. Penentuan bagus atau tidaknya pengambilan keputusan dapat ditentukan dengan metode *gini* ataupun *entropy*. Jadi model akan terus melatih “percabangan” tersebut sampai parameter *minimum sample split* dan *maximum depth* terpenuhi.

2. Bandingkan hasil evaluasi model from scratch dan *library*, bagaimana hasil perbandingannya? Jika ada perbedaan, jelaskan alasannya!

Akurasi [holdout 80-20]	
<i>Scratch</i>	<i>Library</i>
0.672	0.683
Akurasi [k-fold 5]	
<i>Scratch</i>	<i>Library</i>
0.720	0.703
Waktu Eksekusi	
<i>Scratch</i>	<i>Library</i>
144 detik	0.032 detik

3. Jelaskan *improvement* apa saja yang bisa Anda lakukan untuk mencapai hasil yang lebih baik dibandingkan dengan hasil yang Anda punya saat ini! *Improvement* yang dimaksud tidak terbatas pada bagaimana algoritma diimplementasikan, namun juga mencakup tahap sebelum *modeling and validation*.

⇒ Feature selection : memilih fitur-fitur yang relevan dapat membantu meningkatkan kinerja model.

⇒ Hyperparameter tuning : mengatur hyperparameter seperti *minimum sample split* dan *maximum depth* dapat mempengaruhi hasil yang dihasilkan.

⇒ Algoritma : melakukan “pemangkasan” pada *node* atau cabang yang tidak memberi kontribusi signifikan terhadap akurasi model.