BAB II

LANDASAN TEORI

2.1 Artificial Intelligence

Menurut Kok dkk (Kok, et al., 2009), definisi dan arti dari kata *artificial intelligence* adalah sebuah topik yang banyak didiskusikan dan banyak menyebabkan kekeliruan. Satu kamus saja memiliki beberapa definisi dari *artificial intelligence*, salah satunya dari The New International Webster's Comprehensive Dictionary of the English Language (Smith, 2003) adalah suatu bidang studi pada ilmu komputer yang berhubungan dengan pengembangan kemampuan komputer untuk melakukan proses berpikir seperti manusia seperti belajar, berpikir logis, dan koreksi diri.

Definisi dan arti dari *artificial intelligence* berubah-ubah seiringnya waktu dan perkembangan yang pesat. Definisi dan arti terkini membicarakan tentang meniru kecerdasan dan perilaku manusia. Dibandingkan hanya melihat definisi *artificial intelligence* secara umum, *artificial intelligence* dapat dikategorikan menjadi empat (4), yaitu:

a. Sistem yang berpikir seperti manusia.

b. Sistem yang bertindak seperti manusia.

c. Sistem yang berpikir secara rasional.

d. Sistem yang bertindak secara rasional.

2.2 Machine Learning

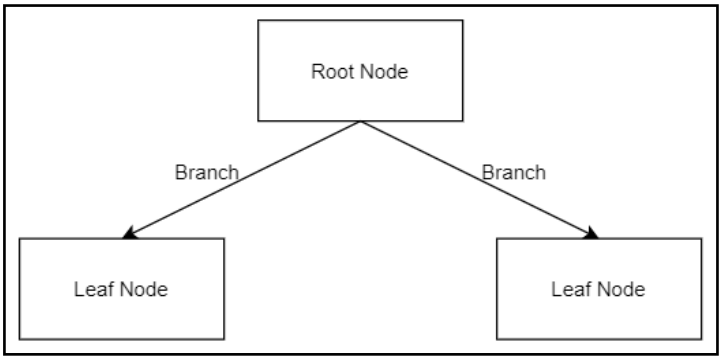
Menurut Saitta (Saitta, 2009), *machine learning* mencari jalan untuk komputer dapat belajar dari data atau pengalaman masa lalu, tujuan utamanya adalah untuk membantu manusia dalam membangun program-program yang tidak dapat dibangun secara langsung, dan program-program yang belajar dan berkembang dari pengalaman. Tujuan lain dari *machine learning* adalah untuk menyediakan model komputasi untuk pembelajaran manusia dan mendukung studi kognitif pembelajaran.

2.3 Supervised Learning/Unsupervised Learning

Dalam *supervised learning*, *agent* memperhatikan adanya *input-output* pada data dan mempelajari suatu fungsi untuk memetakan dari *input* menjadi *output*. Sedangkan dalam *unsupervised learning*, *agent* mempelajari pola dari *input* walaupun tidak disediakan umpan balik. Tugas paling umum dari *unsupervised learning* adalah *clustering*, mendeteksi klaster-klaster yang berpotensi berguna dari input (Russell & Norvig, 2010).

2.4 Decision Tree Learning

Decision tree learning adalah salah satu bentuk pembelajaran mesin paling sederhana dan paling sukses. *Decision tree* merepresentasikan sebuah fungsi yang menerima *input-input* dan mengeluarkan *output* sebuah keputusan. Tipe *input* dan *output* dapat berjenis diskret atau kontinu (Russell & Norvig, 2010).



Gambar 2.1 Elemen-elemen *decision tree*

Sebuah *decision tree* akan sampai di keputusan akhir dengan melakukan rangkaian pengujian. Setiap *node* dalam *tree* berkorespondensi dengan sebuah pengujian nilai dari salah satu atribut *input*. Setiap label cabang (*branch*) dari *node* adalah nilai dari atribut yang diuji. Setiap daun *node* (*leaf*) adalah nilai *output* dari rangkaian pengujian (Russell & Norvig, 2010).

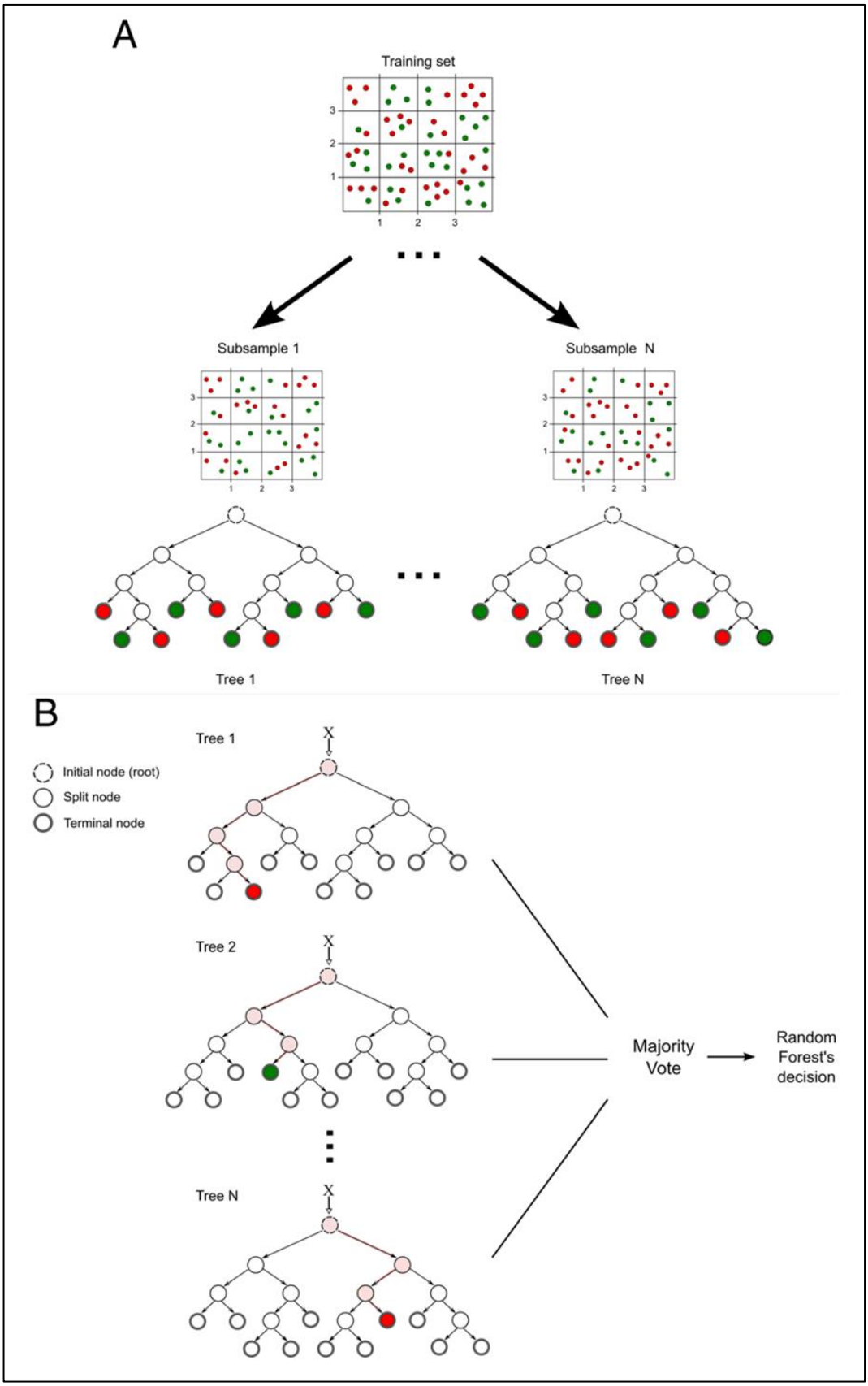
2.5 Bootstrap Aggregation

Bootstrap aggregation atau *bagging* adalah suatu prosedur yang bertujuan untuk mengurangi variasi dalam metode-metode pembelajaran statistik. Umumnya untuk mengurangi variasi dan meningkatkan akurasi prediksi dari suatu metode pembelajaran statistik adalah mengambil banyak set pengujian, membangun model prediksi terpisah dari masing-masing set pengujian, dan merata-rata hasil prediksi. Namun ini tidak praktis karena umumnya kita tidak memiliki akses ke beberapa set pengujian. Melainkan, kita bisa lakukan *bootstrap* dengan mengambil sampel berulang-ulang kali dari suatu set pengujian yang sama. Dengan melalui pendekatan ini kita menghasilkan *N dataset* pengujian berbeda-beda yang telah di-*bootstrap*. Lalu melatih metode di setiap set pengujian yang telah di-*bootstrap* dan dirata-rata dari semua hasil prediksi (James, et al., 2017).

2.6 Random Forest

Random Forest adalah sekumpulan *tree predictors* sederhana yang tiap *tree* menghasilkan *output* dari suatu set nilai *predictor* yang diberikan sebagai *input*. *Random Forest* bisa digunakan untuk permasalahan klasifikasi dan regresi (Gupta, et al., 2017). Untuk permasalahan klasifikasi, jumlah *predictor* standar adalah $\sqrt{p}$ dengan minimum jumlah *node* adalah satu (1), dan untuk permasalahan regresi, jumlah *predictor* standar adalah $p/3$ dengan minimum jumlah *node* adalah lima (5), dimana p adalah jumlah total *predictor* sebelum dilakukan *bagging*. Dalam praktisnya nilai untuk parameter-parameter diatas akan bergantung pada permasalahan dan parameter tersebut diperlakukan sebagai parameter penyetelan (Hastie, et al., 2017).

Gambar 2.2 merupakan ilustrasi cara kerja Random Forest. *Dataset* dipisahkan menjadi *dataset* pelatihan dan *dataset* pengujian. Dibuat sebanyak N *subsample dataset* yang lalu akan dibuat *tree-tree* berdasarkan *subsample dataset* tersebut. Hasil dari semua *tree* tersebut akan dikumpulkan dan suara terbanyak akan dijadikan hasil dari Random Forest untuk permasalahan klasifikasi. Untuk permasalahan regresi, hasil dari semua *tree* akan dirata-rata untuk menjadi hasil dari Random Forest.



Gambar 2.2 Cara kerja Random Forest (Machado, et al., 2015)