**Judul : Classification Based on Decision Tree Algorithm for Machine Learning**

**Sumber : Journal of Aplplied Science and Technology Trends**

**Penulis : Bahzad Taha Jijo, Adnan Mohsin Abdulazeez**

Decision tree adalah teknik pembelajaran **Supervised** yang dapat digunakan untuk masalah klasifikasi dan regresi. Tetapi lebih umum digunakan untuk menyelesaikan masalah klasifikasi.

* **Introduce Decision Tree**

Decision tree / pohon keputusan merupakan metode klasifikasi yang menerapkan stuktur pohon. Konsep dari decision tree adalah mengubah tumpukan data menjadi sebuah pohon keputusan yang mereprentasikan aturan-aturan dari sebuah keputusan.

Bentuk dari pohon keputusan sbb. Lihat gambar

Beberapa terminology :

1. Root node : atau biasa disebut sebagai root atau akar. Root akan menempati posisi paling atas. Node akar berada diawal pohon keputusan. Ini mewakili seluruh populasi yang dianalisis.
2. Splitting : proses membagi node menjadi sub-node apabila node tsb masih dapat dipecah menjadi node lebih lanjut.
3. Internal node : sejumlah node yang berada diantara root dan leaf. Biasa dikenal dengan internal node/decision node. Jika dilihat pada slide ciri ciri dari internal node adalah hanya memiliki 1 node yang masuk tetapi bisa memiliki 1 / 2 node yang keluar. Setiap decision node berhubungan dengan suatu pertanyaan atau pengujian.
4. Leaf node/terminal node : sejumlah node yang posisinya paling bawah. Leaf node dari stuktur pohon tidak memiliki lagi pecahan sub node atau node yang keluar. Setiap leaf node dari decision tree akan mereprensentasikan prediksi class yang akan dihasilkan dari struktur tree tersebut. Berarti ini sudah menunjukkan sebuah label atau hasil dari klasifikasi
5. Pruning : ketika kita menghapus sub-node dari sebuah node, proses ini disebut pruning. Ini adalah kebalikan dari proses splitting yang digunakan untuk mencegah overfitting.

* **Types of Decision Tree Algorithm**

1. CART
2. C4.5
3. CHAID
4. QUEST

* **Creating a Decision Tree from data**
* **Calculate Information Gain**
* Information gain : selisih antara entropy dataset awal dengan rata-rata terboboti dari masing-masing bagian.
* Information gain bisa diperoleh dengan menselisihkan nilai gini impurity sebelum proses splitting dengan nilai rata-rata gini impurity setelah proses splitting.
* **Peranan information gain** :

Gini impurity (pengukuran ketidakmurnian). Gini impurity memiliki jangkauan nilai 0 dan1. 0 mengindikasikan nilai murni yang sempurna. 1 nilai paling tidak murni. P (probability).

* **Constuct Decision Tree**

1. Split root node : Pilih rele keputusan yang memiliki perolehan informasi terbesar untuk membagi simpul saat ini.
2. Split child node : Hanya pertimbangkan keputusansayapada aturan yang tidak pernah dipilih sebelumnya di cabang saat ini.
3. Split child node : Stop splitting if there are no decision rules left or the grup impurity = 0.

* **Benefits and Drawbacks**

Benefits :

1. Sederhana untuk dipahami.
2. Cepat diterjemahkan ke satu set prinsip untuk produksi.
3. Dapat mengklasifikasikan hasil kategorikal dan numerik, tetapi atribut yang dihasilkan harus kategoris.
4. Tidak ada hipotesis apriori yang diambil dengan mempertimbangkan kebaikan hasil.

Drawbacks :

1. Mekanisme pengambilan keputusan yang optimal dapat dicegah dan keputusan yang salah dapat mengikuti.
2. Ada banyak lapisan dikeputusanpohon, yang membuatnya menarik.
3. Untuk sampel pelatihan yang lebih banyak, kompleksitas perhitungan pohon keputusan dapat meningkat.