

BAB II TINJAUAN PUSTAKA

2.1 Penelitian Terdahulu

Tabel 2.1 Penelitian Terdahulu

Nama Penulis	Judul Penelitian	Hasil Penelitian
Indra Mukti P. serta Subiyanto	Sistem Rekomendasi Penjurusan Sekolah Menengah Kejuruan dengan Algoritma C4.5 Jurnal Kependidikan, Volume 1, Nomor 1, Juni 2017, Halaman 139-149	Disini peneliti memiliki sampel yang berjumlah 48 data siswa diambil dari peringkat tiga besar di Kecamatan Batang kemudian proses dengan algoritma C4.5
Rina Kurniasari Dan Azizah Fatmawati	Penerapan Algoritma C4.5 untuk Penjurusan Siswa Sekolah Menengah Atas Jurnal Ilmiah Komputer dan Informatika Vol. 8, No. 1, Maret 2019, ISSN : 2089-9033, eISSN : 2715-7849	Disini penulis mencoba untuk mengklasifikasi dari tiga jurusan dengan perhitungan manual menggunakan excel dan kemudian dijadikan website
Beti Novianti 1) Tedy Rismawan 2) dan Syamsul Bahri 3)	Implementasi Data Mining dengan Algoritma C4.5 untuk Penjurusan Siswa (Studi Kasus: SMA Negeri 1 Pontianak)	Jumlah data yang digunakan peneliti pada kasus ini berjumlah 39 data siswa dengan dua pengklasifikasian

	Jurnal Coding, Sistem Komputer Untan Volume 04, No.3 (2016), hal 75-84	penjurusan dengan hasil 8 siswa masuk IPS dan 31 siswa masuk IPA
--	--	--

2.2 Data Mining

Tidak asing lagi Data mining secara istilah adalah ekstraksi pada suatu database sehingga ditemukan sebuah informasi maupun pengetahuan yang baru. Dalam proses ekstraksi dan pengidentifikasian dari sebuah database diperlukan machine learning dan lain-lain untuk mendapatkan informasi. KDD yakni pencarian keteraturan penyebutan lain daripada data mining itu sendiri, hubungan antara set data yang terkumpul (Indra dan Subiyanto, 2017)

Pembaharuan sebuah pengetahuan bisa didapatkan dari data lama yang dianggap tidak ada gunanya melalui pengolahan data maka bisa menjadi sebuah landasan untuk menentukan sebuah keputusan yang bijak dan kuat. Berikut ini fungsi data mining

- a. Deskriptif
- b. Prediktif
- c. Asosiatif
- d. Estimatif
- e. Pengelompokan

2.3 Algoritma C4.5

Adalah sebuah algoritma menitik beratkan pada pembuatan sebuah pohon keputusan yang berlandaskan data. Algoritma C4.5 sendiri merupakan hasil algoritma ID3 yang telah dimodifikasi. Secara sederhana algoritma C4.5 merubah dari kumpulan data lalu dijadikan pohon keputusan untuk menghasilkan sebuah aturan-aturan yang baru

Ada empat tahapan secara umum dalam penerapan algoritma C4.5 dalam memperoleh aturan keputusan yakni :

- a. Pemilihan atribut untuk dijadikan root atau akar
- b. Pembuatan cabang berdasarkan nilai-nilai yang ada
- c. Pembagian kasus untuk tiap-tiap cabang

- d. Lakukan pengulangan untuk setiap cabang sehingga semua kelas dari tiap-tiap cabang menjadi sama

Adapun tahap-tahap khusus algoritma C4.5 memakai kriteria *split* yang kemudian mengalami perubahan dalam tahap penentu *split* untuk penyebutannya yakni *Gain Ratio*. Yang dimaksud dengan *split* adalah tahapan utama dalam pembuatan pohon keputusan

Berikut ini merupakan tahap-tahap khusus dalam penggunaan algoritma C4.5 yang telah dimodifikasi :

- a. Mencari nilai *Entropy*
- b. Menghitung *Gain Ratio* masing-masing atribut yang ada
- c. Kemudian *Gain* tertinggi dijadikan akar dan *Gain Ratio* tertinggi berikutnya yang memiliki nilai lebih kecil dari pada akar akan dijadikan sebagai cabangnya
- d. Mengulangi tahap ke-2 dan ke-3 tanpa atribut yang sudah dijadikan root
- e. *Gain Ratio* tertinggi dari atribut akan dijadikan cabang
- f. Tahap terakhir mengulangi lagi tahapan 4 dan 5 sehingga *Gain Ratio* yang didapatkan bernilai 0 dari atribut-atribut yang tersisa

Bisa menggunakan rumus dibawah ini dalam menemukan nilai *Entropy* pada algoritma ini :

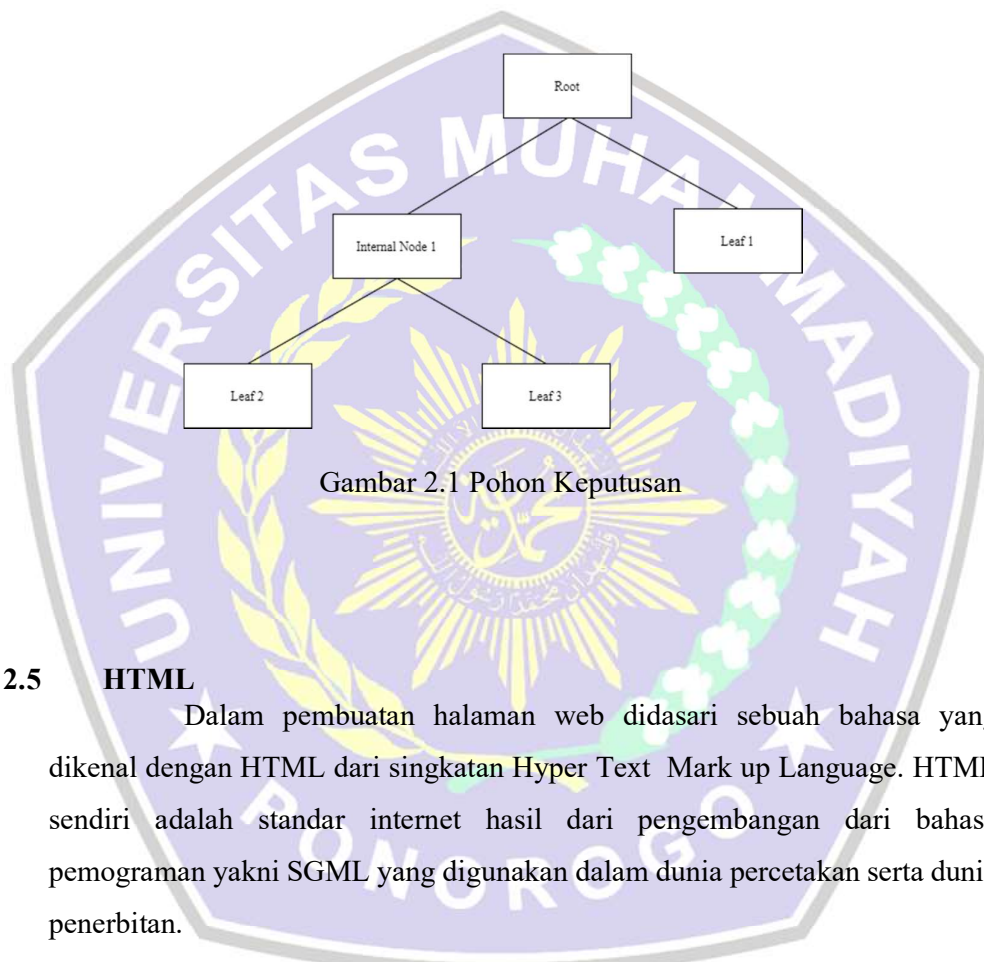
$$Entropy(S) = \sum_{i=1}^n p_i * \log_2 p_i \quad \dots(2.1)$$

Adapun untuk mendapatkan nilai *information gain* kita menggunakan rumus seperti dibawah ini :

$$Gain(S,A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \quad \dots(2.2)$$

2.4 **Pohon Keputusan**

Algoritma ini adalah pengklasifikasian dengan struktur sederhana dan sering kali digunakan untuk penelitian selain itu juga mudah untuk diterapkan yakni Pohon Keputusan. Sering kali disebut dengan pohon terbalik karena bentuknya menyerupai pohon terbalik dengan filosofi akar terletak diatas sedangkan daunnya berada dibawah



Gambar 2.1 Pohon Keputusan

2.5 **HTML**

Dalam pembuatan halaman web didasari sebuah bahasa yang dikenal dengan HTML dari singkatan Hyper Text Mark up Language. HTML sendiri adalah standar internet hasil dari pengembangan dari bahasa pemograman yakni SGML yang digunakan dalam dunia percetakan serta dunia penerbitan.

Sebagai penyusun kerangka serta struktur dari sebuah halaman web HTML harus dikombinasikan dengan bahasa CSS untuk mendapatkan tampilan website yang menarik serta dikolaborasikan dengan JavaScript untuk fungsi interaksi dengan penggunanya.

2.6 MYSQL

Untuk pembuatan aplikasi maupun website dinamis yang mana aplikasi itu mempunyai banyak model data serta informasi meliputi nama user, kata sandi dan banyak lagi, sangatlah membutuhkan database salah satunya adalah MySQL

Dalam pembuatan aplikasi berbasis website sangatlah diperlukan yang namanya DBMS yang mana dalam perintahnya menggunakan SQL. Ada dua pembagian dari lisensi MySQL diantaranya Free Software dan Shareware. Untuk menghubungkan aplikasi atau website dengan database server yang berisi tabel dan kolom. MySQL ini menggunakan metode relational database dalam sistem pengambilan datanya

Ada beberapa fungsi dari MySQL yang pertama yakni membuat dan mengelola database yang ada di server, database tersebut bisa memuat bermacam-macam informasi dengan menggunakan bahasa SQL. Dan yang kedua fungsi lain yang bisa dilakukan MySQL adalah memudahkan user dalam mengambil kembali data yang berisi informasi berbentuk teks, sehingga pihak publik atau personal dapat diakses kembali dengan mudah menggunakan perintah SQL

