

MEMPREDIKSI TINGKAT PEMINAT EKSTRAKURIKULER PADA SISWA SMK ANALISIS KESEHATAN ABDURRAB MENGGUNAKAN ALGORITMA C4.5 (STUDI KASUS: SMK ANALIS KESEHATAN ABDURRAB)

¹⁾Luluk Elvitaria, ²⁾Muhammad Havenda

Program Studi Teknik Informatika, Fakultas Teknik, Universitas Abdurrah
Jalan Riau Ujung No. 73 Pekanbaru – Riau 28292
E-Mail : luluk@univrab.ac.id

ABSTRAK

Kegiatan ekstrakurikuler adalah sebuah kegiatan tambahan di sekolah, yang di mana melalui kegiatan ini, siswa dapat menambah atau menggali keterampilan siswa-siswi dalam upaya pembinaan diri. Salah satu dari kegiatan ekstrakurikuler adalah ekstrakurikuler bahasa asing, meliputi 5 bahasa yaitu Arab, Inggris, Jerman, Mandarin, Jepang. Dalam mengetahui ketertarikan siswa terhadap kegiatan ekstrakurikuler, maka dilakukan penelitian mengenai tingkat peminat kegiatan ekstrakurikuler yaitu bahasa asing pada siswa SMK Analisis Kesehatan Abdurrah. Dalam memprediksi tingkat peminat bahasa asing dengan proses *data mining* menggunakan metode Algoritma C4.5. Algoritma C4.5 merupakan kelompok dari Algoritma *Decision Tree*. Dari penelitian ini, pihak sekolah dapat mengetahui sejauh mana tingkat peminat bahasa asing pada siswa-siswi dan sekolah dapat meningkatkan kegiatan ekstrakurikuler serta siswa-siswi dapat mengembangkan minatnya terhadap bahasa asing sesuai keinginannya.

Kata kunci: *Kegiatan ekstrakurikuler, bahasa asing, siswa-siswi, algoritma C4.5, decision tree*

ABSTRACT

Extracurricular activities are an additional activities at the school, which is where through these activities, students can add or digging skills of students in the self development efforts. One of extracurricular activities is extracurricular other language, includes 5 languages are Arabic, English, German, Mandarin, Japanese. In know the interest of students to extracurricular activities, then do research on the level of interest in extracurricular activities is a foreign language at vocation high school students of Health Analyst Abdurrah. In predicting the level of interest in other language with the data mining process using C4.5 Algorithm. Algorithm C4.5 is a group of the Decision Tree algorithm. From this research, the school can determine the extent of interest in a foreign language at students and schools can improve extracurricular activities and students can develop his interest in foreign languages as he wishes.

Keywords: *Extracurricular activities, other languages, students, C4.5 algorithm, descision tree*

PENDAHULUAN

Kegiatan ekstrakurikuler adalah sebuah kegiatan tambahan di sekolah. Melalui kegiatan ini, siswa dapat menambah atau menggali keterampilan siswa-siswi dalam upaya pembinaan diri. Sebelum kegiatan ekstrakurikuler ditetapkan dan disosialisasikan oleh pihak sekolah kepada siswa, pihak sekolah terlebih dahulu mengetahui sejauh mana ketertarikan siswa terhadap kegiatan ekstrakurikuler yang akan diadakan. Salah satu dari kegiatan

ekstrakurikuler adalah ekstrakurikuler bahasa asing, meliputi 5 bahasa yaitu Arab, Inggris, Jerman, Mandarin, Jepang. Bahasa adalah sebuah sistem bunyi yang arbitrer yang digunakan oleh masyarakat untuk tujuan komunikasi.

Dalam mengetahui ketertarikan siswa terhadap kegiatan ekstrakurikuler, maka dilakukan penelitian mengenai tingkat peminat kegiatan ekstrakurikuler yaitu bahasa asing pada siswa SMK Analisis Kesehatan Abdurrah. Dalam memprediksi tingkat peminat bahasa asing dengan proses

data mining menggunakan metode Algoritma C45. Algoritma C45 merupakan kelompok dari Algoritma *Decision Tree*.

Menurut David Hartanto Kamagi & Seng Hansun (2014), bahwa algoritma C4.5 dapat diimplementasikan untuk memprediksi tingkat kelulusan mahasiswa dengan empat kategori yaitu lulus cepat, lulus tepat, lulus terlambat dan drop out. Atribut yang paling berpengaruh dalam hasil prediksi adalah IPS semester enam.

Dari penelitian ini, pihak sekolah dapat mengetahui sejauh mana tingkat peminat bahasa asing pada siswa-siswi dan sekolah dapat meningkatkan kegiatan ekstrakurikuler serta siswa-siswi dapat mengembangkan minatnya terhadap bahasa asing sesuai keinginannya.

Rumusan Masalah

Berdasarkan pada latar belakang masalah yang telah dijelaskan sebelumnya, maka permasalahan pokok yang dibahas pada tugas akhir ini adalah :

- a. Bagaimana memprediksi tingkat peminat bahasa asing pada siswa-siswi SMK Analis Kesehatan Abdurrah?
- b. Variabel apa yang diperlukan dalam memprediksi tingkat peminat bahasa asing pada siswa-siswa SMK Analis Kesehatan Abdurrah?
- c. Bagaimana menerapkan metode algoritma C4.5 dalam memprediksi tingkat peminat bahasa asing pada siswa-siswi SMK Analis Kesehatan Abdurrah?

Batasan Masalah

Ruang lingkup yang telah didapat dari penelitian ini adalah:

- a. Kegiatan ekstrakurikuler yaitu bahasa asing meliputi Arab, Inggris, Jerman, Mandarin, Jepang.
- b. Data Mining yaitu metode Algoritma C4.5

c. Variabel dalam memprediksi tingkat peminat bahasa asing adalah hasil dari kuesioner

d. Penelitian dilakukan pada siswa-siswi SMK Analis Kesehatan Abdurrah

Tujuan

Adapun tujuan penelitian yang akan dicapai dalam penulisan ini adalah sebagai berikut :

- a. Menganalisa dalam memprediksi tingkat peminat bahasa asing pada siswa-siswi SMK Analis Kesehatan Abdurrah.
- b. Menerapkan algoritma C4.5 dalam memprediksi tingkat peminat bahasa asing pada siswa-siswi SMK Analis Kesehatan.
- a. Mengetahui tingkat peminat bahasa asing dari penentuan pohon keputusan dengan menghitung nilai *entropy* dan *gain*.

Manfaat

Manfaat yang dihasilkan dalam penelitian ini adalah:

- a. Dapat membantu dalam mengetahui tingkat peminat bahasa asing.
- b. Sebagai penunjang sekolah dalam meningkatkan peminat bahasa asing.
- c. Dapat menyelesaikan masalah dalam memprediksi tingkat peminat bahasa asing pada siswa-siswi SMK Abdurrah.

Data Mining

Menurut David dan Seng (2014) *Data Mining* didefinisikan sebagai sebuah proses untuk menemukan hubungan, pola dan tren baru yang bermakna dengan menyaring data yang sangat besar, yang tersimpan dalam penyimpanan, menggunakan teknik pengenalan pola seperti teknik Statistik dan Matematika.

Menurut Rizky dan Nita (2013) *Data Mining* adalah serangkaian proses untuk menggali nilai tambah yang berupa informasi

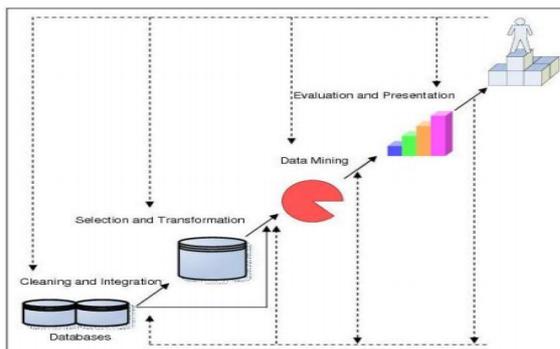
yang selama ini tidak diketahui secara manual dari suatu basis data dengan melakukan proses ekstraksi dan mengenali pola penting dari data yang ada.

Cross-Industry Standard Process for Data Mining (CRISP-DM)

Data mining telah diterapkan di hampir seluruh bidang industri dan pengetahuan. Dengan semakin luasnya penerapan *data mining* tersebut, terdapat keinginan dari sekelompok analis *data mining* yang mewakili DaimlerChrysler, SPSS, dan NCR untuk membuat sebuah model proses *data mining* yang netral terhadap jenis industri, *tools*, dan aplikasi (Indri, 2014).

Proses Data Mining

Menurut (Indri, 2014). *Data Mining* adalah sebuah proses untuk menemukan pola atau pengetahuan yang bermanfaat secara otomatis dari sekumpulan data yang berjumlah banyak, *Data Mining* sering dianggap sebagai bagian dari *Knowledge Discovery in Database (KDD)* yaitu sebuah



proses mencari pengetahuan yang bermanfaat dari data.

Gambar 1. KDD (Sunjana, 2010):

1. Data Selection

Pemilihan (seleksi) data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam KDD dimulai. Data hasil seleksi yang akan

digunakan untuk proses *data mining*, disimpan dalam suatu berkas, terpisah dari basis data operasional.

2. Pre-processing/Cleaning

Sebelum proses *data mining* dapat dilaksanakan, perlu dilakukan proses *cleaning* pada data yang menjadi focus KDD. Proses *cleaning* mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak (tipografi). Selain itu dilakukan proses *enrichment*, yaitu proses “memperkaya” data yang sudah ada dengan data atau informasi lain yang relevan dan diperlakukan untuk KDD, seperti data atau informasi eksternal.

3. Transformation

Coding adalah proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses *data mining*. Proses *coding* dalam KDD merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data.

4. Interpretation/Evaluation

Pola informasi yang dihasilkan dari proses *data mining* perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini merupakan bagian dari proses KDD yang disebut dengan *interpretation*. Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesa yang ada sebelumnya.

Pengolahan Data Mining

Pengolahan data mining terdiri dari beberapa metode pengolahan, yaitu ((Indri, 2014). *Predictive modelling* yang merupakan pengolahan data mining dengan melakukan prediksi atau peramalan. Tujuan metode ini untuk membangun model prediksi suatu nilai yang mempunyai ciri-ciri tertentu. Contoh algoritmanya *Linear Regression*, *Neural*

Network, *Support Vector Machine*, dan lain-lain.

1. *Association* (Asosiasi) merupakan teknik dalam *data mining* yang mempelajari hubungan antar data. Contoh penggunaannya seperti untuk menganalisis perilaku mahasiswa yang datang terlambat. Contohnya jika mahasiswa memiliki jadwal dengan dosen A dan B, maka mahasiswa akan datang terlambat. Contoh algoritmanya *FP-Growth*, *A Priori*, dan lain-lain.
2. *Clustering* (Klastering) atau pengelompokan merupakan teknik untuk mengelompokkan data ke dalam suatu kelompok tertentu. Contoh algoritmanya *K-Means*, *K-Medoids*, *Self-Organisation Map* (SOM), *Fuzzy C-Means*, dan lain-lain. Contoh untuk *clustering*: Terdapat lima pulau di Indonesia: Sumatera, Kalimantan, Jawa, Sulawesi dan Papua. Maka lima pulau tersebut dijadikan tiga klaster berdasarkan waktunya: Waktu Indonesia Barat (Sumatera, Kalimantan dan Jawa), Waktu Indonesia Tengah (Sulawesi) dan Waktu Indonesia Timur (Papua).
3. *Classification* merupakan teknik mengklasifikasikan data. Perbedaannya dengan metode *clustering* terletak pada data, di mana pada *clustering* variabel dependen tidak ada, sedangkan pada *classification* diharuskan ada variabel dependen. Contoh algoritma yang menggunakan metode ini ID3 dan *K Nearest Neighbors*.

Pengelompokan Data Mining

Data mining dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dikakukan, yaitu (Jefri & Kusri, 2013):

1. Prediksi

Prediksi hampir sama dengan klasifikasi dan estimasi, kecuali bahwa dalam prediksi nilai hasil mungkin akan ada di masa mendatang. Contoh prediksi dalam penelitian:

- a. Prediksi harga gula dalam tiga dekade yang akan datang
- b. Prediksi keadaan cuaca pada suatu tempat apakah akan terang, mendung, hujan dan sebagainya.

Beberapa metode atau teknik yang digunakan dalam klasifikasi dan estimasi dapat juga digunakan (untuk keadaan yang tepat) untuk memprediksi.

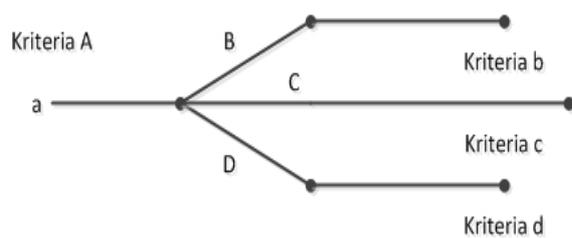
2. Klasifikasi

Didalam pengklasifikasian terdapat target variabel katagori. Sebagai contoh, penggolongan pendapatan dapat dipisahkan dalam tiga katagori, yaitu pendapatan tinggi, pendapatan sedang, dan pendapatan rendah. Contoh lain klasifikasi dalam penelitian adalah:

- a. Menentukan apakah suatu transaksi kartu kredit merupakan transaksi yang curang atau tidak.
- b. Mendiagnosis penyakit seorang pasien untuk mendapatkan termasuk katagori penyakit apa.

Pohon Keputusan (Decision Tree)

Pohon keputusan adalah model prediksi menggunakan struktur pohon atau struktur berhirarki. *Decision Tree* adalah *flow-chart* seperti struktur *tree*, dimana tiap *internal node* menunjukkan sebuah *test* pada sebuah atribut, tiap cabang menunjukkan hasil dari *test* dan *leaf node* menunjukkan *class-class* atau *class distribution* (Selvia. et al, 2014).



Gambar 2. Susunan Pohon Keputusan

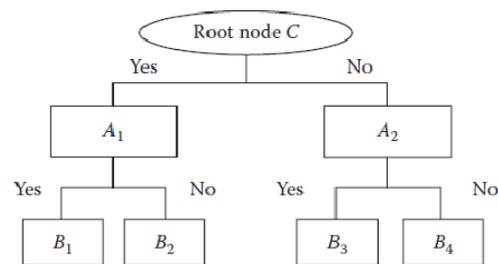
Metode pohon keputusan mengubah fakta yang sangat besar menjadi pohon keputusan yang merepresentasikan aturan. Aturan dapat dengan mudah dipahami dengan bahasa alami, juga dapat diekspresikan dalam bentuk bahasa basis data seperti *Structured Query Language* untuk mencari *record* pada kategori tertentu (David & Seng, 2014).

Pada *decision tree* terdapat 3 jenis *node*, yaitu (Anik, 2013):

1. *Root node*, merupakan *node* paling atas, pada *node* ini tidak ada *input* dan bisa tidak mempunyai *output* atau mempunyai *output* lebih dari satu.
2. *Internal node*, merupakan *node* percabangan, pada *node* ini hanya terdapat satu *input* dan mempunyai *output* minimal dua.
3. *Leaf node* atau *terminal node*, merupakan *node* akhir, pada *node* ini hanya terdapat satu *input* dan tidak mempunyai *output*.

1. Struktur *Decision Tree*

Menurut Dua dan Xian, 2011, seperti ditunjukkan dalam Gambar 2.4, *decision tree* tergantung pada aturan *if-then*, tetapi tidak membutuhkan parameter dan metrik. Struktur sederhana dan dapat ditafsirkan memungkinkan *decision tree* untuk memecahkan masalah atribut *multi type*. *Decision tree* juga dapat mengelola nilai-nilai yang hilang atau data *noise* (Anik, 2013).



Gambar 3. Pohon Keputusan

Tahapan *Decision Tree*

Ada beberapa tahap dalam membuat sebuah pohon keputusan yaitu (Mila & Dedi, 2015):

1. Menyiapkan data *training* yang sudah dikelompokkan ke dalam kelas-kelas tertentu.
2. Menentukan akar dari pohon. Akar akan diambil dari atribut yang terpilih, yaitu dengan cara menghitung nilai *gain* dari masing – masing atribut. Nilai *gain* yang tertinggi akan menjadi akar pertama. Sebelum menghitung nilai *gain* dari atribut, hitung nilai *entropy*.

Algoritma C4.5

Algoritma C4.5 merupakan generasi baru dari algoritma ID3 yang dikembangkan oleh J.Ross Quinlan pada tahun 1983 (Windy. *et al*, 2014). Algoritma C4.5 diperkenalkan oleh Quinlan (1996) sebagai versi perbaikan dari ID3 (Eko, 2014). Sebelum membahas algoritma C4.5 perlu dijelaskan terlebih dahulu algoritma ID3 karena C4.5 adalah ekstensi dari algoritma *decision tree* ID3 (Sunjana, 2010).

Pembuatan pohon keputusan menggunakan algoritma C4.5 yang merupakan pengembangan dari algoritma ID3, dimana pengembangan dilakukan dalam hal mengatasi *missing data*, data *continue*, *pruning*. Secara umum, algoritma C4.5 untuk membangun pohon keputusan dimulai dari pemilihan atribut sebagai akar, membuat cabang untuk tiap-tiap nilai, membagi kasus

dalam cabang dan mengulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama (Rizky & Nita, 2013).

Algoritma C4.5 merupakan kelompok algoritma *Decision Tree*. Algoritma ini mempunyai input berupa *training samples* dan *samples*. *Training samples* berupa data contoh yang akan digunakan untuk membangun sebuah *tree* yang telah diuji kebenarannya. Sedangkan *samples* merupakan *field – field* data yang nantinya akan digunakan sebagai parameter dalam melakukan klasifikasi data (Selvia. *et al*, 2014).

Secara umum Algoritma C4.5 untuk membangun pohon keputusan adalah sebagai berikut (Jefri & Kusri, 2013):

- a. Pilih atribut sebagai akar.
- b. Buat cabang untuk masing-masing nilai.
- c. Bagi kasus dalam cabang.
- d. Ulangi proses untuk masing-masing cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Prinsip Kerja Algoritma C4.5

Pada tahap pembelajaran algoritma C4.5 memiliki 2 prinsip kerja yaitu (Selvia. *et al*, 2014):

1. Pembuatan pohon keputusan. Tujuan dari algoritma penginduksi pohon keputusan adalah mengkonstruksi struktur data pohon yang dapat digunakan untuk memprediksi kelas dari sebuah kasus atau *record* baru yang belum memiliki kelas. C4.5 melakukan konstruksi pohon keputusan dengan metode *divide and conquer*. Pada awalnya hanya dibuat *node* akar dengan menerapkan algoritma *divide and conquer*. Algoritma ini memilih pemecahan kasus – kasus yang terbaik dengan menghitung dan membandingkan *gain ratio*, kemudian *node-node* yang

terbentuk di level berikutnya, algoritma *divide and conquer* akan diterapkan lagi sampai terbentuk daun-daun.

2. Pembuatan aturan-aturan (*rule set*). Aturan-aturan yang terbentuk dari pohon keputusan akan membentuk suatu kondisi dalam bentuk *if-then*. Aturan-aturan ini didapat dengan cara menelusuri pohon keputusan dari akar sampai daun. Setiap *node* dan syarat percabangan akan membentuk suatu kondisi atau suatu *if*, sedangkan untuk nilai-nilai yang terdapat pada daun akan membentuk suatu hasil atau suatu *then*.

Tahapan Algoritma Decision Tree C4.5

Ada beberapa tahap dalam membuat sebuah pohon keputusan dengan algoritma C4.5, yaitu (Prabowo. *at al*, 2015):

1. Menyiapkan *data training*. *Data training* biasanya diambil dari data histori yang pernah terjadi sebelumnya dan sudah dikelompokkan ke dalam kelas-kelas tertentu.
2. Menentukan akar dari pohon. Akar akan diambil dari atribut yang terpilih, dengan cara menghitung nilai *gain* dari masing-masing atribut, nilai *gain* yang paling tinggi yang akan menjadi akar pertama. Sebelum menghitung nilai *gain* dari atribut, hitung dahulu nilai *entropy*. *Entropy* adalah ukuran dari teori informasi yang dapat mengetahui karakteristik dari impurity, dan homogeneity dari kumpulan data (Windy. *et al*, 2014). Untuk menghitung nilai *entropy* digunakan rumus:

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

Keterangan:

- S : Himpunan kasus
n : Jumlah partisi S
p_i : Proporsi S_i terhadap S

3. Kemudian hitung nilai *gain*. Gain adalah ukuran efektivitas suatu atribut dalam mengklasifikasikan data (Windy. *et al*, 2014). Hitungan nilai *gain* menggunakan

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \dots\dots\dots (2)$$

rumus:

Keterangan:

S : Himpunan kasus

A : Fitur

n : Jumlah partisi atribut A

|S_i| : Porporasi dari S_i terhadap S

|S| : Jumlah kasus dalam S

4. Ulangi langkah ke-2 hingga semua *record* terpartisi
5. Proses partisi pohon keputusan akan berhenti saat
 - a. Semua *record* dalam simpul N mendapat kelas yang sama.
 - b. Tidak ada atribut di dalam *record* yang dipartisi lagi.
 - c. Tidak ada *record* di dalam cabang kosong.

2. RapidMiner

RapidMiner merupakan perangkat lunak yang bersifat terbuka (open source). *RapidMiner* adalah sebuah solusi untuk melakukan analisis terhadap data mining, text mining dan analisis prediksi. *RapidMiner* menggunakan berbagai teknik deskriptif dan prediksi dalam memberikan wawasan kepada pengguna sehingga dapat membuat keputusan yang paling baik. *RapidMiner* memiliki kurang lebih 500 operator data mining, termasuk operator untuk input, output, data preprocessing dan visualisasi. *RapidMiner* ditulis dengan menggunakan bahasa java sehingga dapat bekerja di semua sistem operasi.

RapidMiner sebelumnya bernama YALE (Yet Another Learning Environment),

dimana versi awalnya mulai dikembangkan pada tahun 2001 oleh RalfKlinkenberg, Ingo Mierswa, dan Simon Fischer di Artificial Intelligence Unit dari University of Dortmund. *RapidMiner* didistribusikan di bawah lisensi AGPL (*GNU Affero General Public License*) versi 3. *RapidMiner* menyediakan GUI (*Graphic User Interface*) untuk merancang sebuah pipeline analitis. GUI ini akan menghasilkan file XML (*Extensible Markup Language*) yang mendefinisikan proses analitis keinginan pengguna untuk diterapkan ke data.

RapidMiner memiliki beberapa sifat sebagai berikut:

1. Ditulis dengan bahasa pemrograman Java sehingga dapat dijalankan di berbagai sistem operasi.
2. Proses penemuan pengetahuan dimodelkan sebagai operator trees.
3. Representasi XML internal untuk memastikan format standar pertukaran data.
4. Bahasa scripting memungkinkan untuk eksperimen skala besar dan otomatisasi eksperimen.
5. Konsep multi-layer untuk menjamin tampilan data yang efisien dan menjamin penanganan data.
6. Memiliki GUI, command line mode, dan Java API yang dapat dipanggil dari program lain.

3. Ekstrakurikuler

Ekstrakurikuler berasal dari kata “kurikuler”, yang artinya adalah kegiatan belajar yang dilakukan melalui tatap muka yang alokasi waktunya sudah ditentukan dalam susunan program dan diperdalam melalui tugas – tugas. Sedangkan “kurikuler” berasal dari kata “kurikulum”, yaitu sejumlah mata pelajaran atau kuliah di sekolah atau perguruan tinggi, yang harus ditempuh untuk

mencapai tujuan pendidikan, juga keseluruhan pelajaran yang disajikan oleh suatu lembaga pendidikan, singkatnya sesuatu yang direncanakan (Abdul, 2011).

Menurut Departemen Pendidikan Nasional (2005:291) kegiatan ekstrakurikuler merupakan suatu kegiatan yang berada di luar program yang tertulis dalam kurikulum, seperti latihan kepemimpinan dan pembinaan siswa (Rizki, 2013).

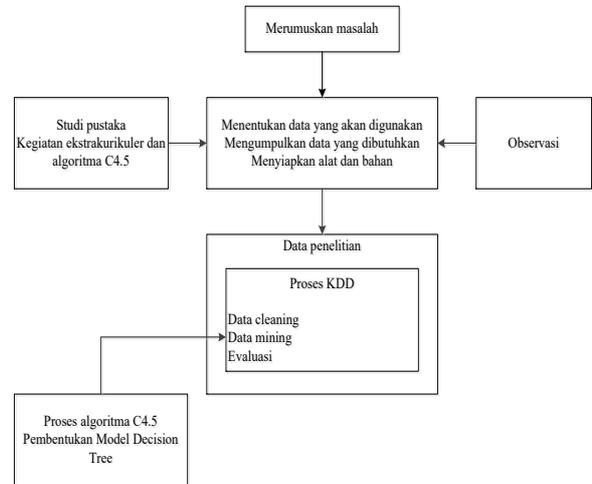
Profil SMK Analis Kesehatan Abdurrah Pekanbaru

Sekolah Menengah Kejuruan (SMK) Abdurrah Pekanbaru Jurusan Analis Kesehatan ini berdiri tahun 1993 atas SK Menkes RI No.HK.00.06.1.3.4869 dan telah mengikuti penilaian akreditasi dari Badan Akreditasi Nasional Sekolah (BAM-S/M) dengan memperoleh jenjang akreditasi Strata A dengan nilai 92 yang merupakan usaha untuk meningkatkan mutu dan kualitas pendidikan di institusi Kesehatan Abdurrah Pekanbaru. Sampai tahun 2015 SMK Abdurrah telah meluluskan siswanya sebanyak 19 angkatan atau 1679 orang siswa atau siswinya. Pada tahun 2015 ini kembali menerima siswa baru angkatan XXII T.A 2015/2016 untuk siswa atau siswi lulusan SMP/MTS/ sederajat.

SMK Abdurrah Pekanbaru Jurusan Analis Kesehatan adalah sekolah lanjutan tingkat atas atau setara SMA yang diperuntukkan bagi tamatan SMP/MTS atau sederajat. Sekolah ini bergerak dibidang kesehatan, dengan lama pendidikan 6 semester atau selama 3 tahun. Penerapan pendidikan dan pengalaman yang dikembangkan dalam bentuk teori di kelas dan praktek di Instansi Rumah Sakit Pemerintah dan Swasta.

METODE

Kerangka Pemikiran



Gambar 4. Kerangka Pemikiran

Di bawah ini merupakan penjelasan dari desain penelitian:

1. Merumuskan masalah, merupakan dasar pemikiran dari penelitian. Rumusan masalah dari penelitian ini:
 - a. Bagaimana memprediksi tingkat peminat bahasa asing pada siswa-siswi SMK Analis Kesehatan Abdurrah?
 - b. Variabel apa yang diperlukan dalam memprediksi tingkat peminat bahasa asing pada siswa-siswa SMK Analis Kesehatan Abdurrah?
 - c. Bagaimana menerapkan metode algoritma C4.5 dalam memprediksi tingkat peminat bahasa asing pada siswa-siswi SMK Analis Kesehatan Abdurrah?
2. Menentukan data yang dibutuhkan

Setelah merumuskan masalah, langkah selanjutnya adalah menentukan data yang dibutuhkan. Data yang dibutuhkan adalah data yang berpengaruh dalam peminat bahasa asing, data tersebut yaitu variabel atau atribut dalam memprediksi peminat bahasa asing dalam bentuk kuesioner dan

- data mengenai SMK Analis Kesehatan Abdurrah
3. Mengumpulkan data yang dibutuhkan
Data yang telah ditentukan pada langkah sebelumnya kemudian dikumpulkan dengan tahap observasi langsung ke Sekolah. Setelah data terkumpul, data tersebut dapat diproses untuk tahap selanjutnya.
 4. Mempersiapkan alat dan bahan penelitian
Mempersiapkan alat dan bahan penelitian. Alat yang dipersiapkan berupa perangkat keras dan perangkat lunak yang menunjang pembuatan sistem. Dan bahan penelitian adalah data-data yang sudah diproses yang kemudian diimplementasikan/ diolah menjadi program. Alat dan bahan ini akan dibahas pada bab selanjutnya.
 5. Studi kepustakaan, merupakan tahapan mengumpulkan data-data baik dari buku bacaan, jurnal, maupun artikel-artikel yang berasal dari internet yang berkaitan dengan algoritma C45 dan juga sistematika penjurusan pada jenjang menengah atas.
 6. Observasi adalah tahapan untuk mengumpulkan data-data penelitian langsung ke Sekolah. Tahap dilakukan dalam observasi ini adalah wawancara kepada pihak sekolah.
 7. Data penelitian didapat dari tahapan-tahapan sebelumnya yaitu tahap merumuskan masalah, menentukan data yang akan digunakan, mengumpulkan data yang dibutuhkan, menyiapkan alat dan bahan, observasi dan studi kepustakaan. Setelah tahapan-tahapan tersebut dilaksanakan maka akan didapatkan data penelitian untuk selanjutnya diproses pada tahap proses Knowledge Discovery in Database (KDD).

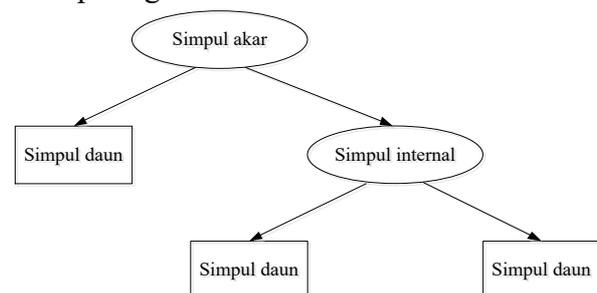
8. Proses Knowledge Discovery in Database KDD

a. Data Cleaning

Proses pembersihan data untuk menghilangkan data tidak lengkap, menghilangkan kesalahan pada data.

b. Data Mining

Tahap ini merupakan tahap untuk menemukan pola pohon keputusan yang sesuai dari data, pada tahap ini algoritma C45 bekerja untuk membentuk pola pohon keputusan. Untuk membentuk model pohon keputusan dari data, tentukan dulu simpul terpilih, yaitu dengan menghitung nilai information gain dari masing-masing atribut. Atribut-atributnya yaitu data penelitian yang sudah dikumpulkan seperti alasan minat bahasa asing, penggunaan bahasa asing, kemampuan bahasa asing, cara pengembangan bahasa asing, kesulitan dalam pengembangan bahasa asing. Simpul yang terpilih nantinya dibagi menjadi simpul akar, simpul internal, dan simpul daun. Dapat digambarkan seperti gambar berikut:



Gambar 5. Contoh Gambar Pohon Keputusan

c. Interpretation/Evaluation

Tahap ini adalah tahap pemeriksaan kesesuaian pohon keputusan yang terbentuk dengan menggunakan algoritma C45 tersebut

HASIL

Analisa Model

Kebutuhan masukan dari sistem adalah berupa atribut yang dimiliki oleh sebuah data nilai atribut dan nilai kemungkinannya yang dibuat kedalam sebuah data tabel. Data tabel yang dimaksud adalah data yang mempunyai minimal dua kolom atribut. Satu kolom sebagai kolom atribut masukan dan satu kolom sebagai kolom atribut target. Berikut adalah data analisis prediksi peminat bahasa asing.

Tabel 1. Tabel Atribut

Atribut	Bagian atribut
Alasan menyukai bahasa asing yang disukai	Mudah dipahami dan dipelajari
	Mengikuti perkembangan zaman
	Pelajaran sekolah
	Lebih mudah ke negara lain
	Ingin menguasai Komunikasi
Menggunakan bahasa asing	Karena bahasa dunia
	Ya
Kemampuan bahasa asing	Tidak
	Aktif
Cara mengembangkan bahasa asing	Pasif
	Les privat
Kesulitan belajar bahasa asing	Belajar sendiri
	Mengikuti kursus
Kesulitan belajar bahasa asing	Ya
	Tidak

Penerapan Algoritma C4.5 Data hasil *cleaning* selanjutnya dianalisa untuk menghasilkan sebuah pohon keputusan dengan menggunakan algoritma C4.5, secara umum algoritma C4.5 untuk membangun pohon keputusan adalah sebagai berikut:

1. Perhitungan Entropy dan Gain

2. Pemilihan Gain tertinggi sebagai akar (Node)
3. Ulangi proses perhitungan Entropy dan Gain untuk mencari cabang sampai semua kasus pada cabang memiliki kelas yang sama yaitu pada saat semua variabel telah menjadi bagian dari pohon keputusan atau masing-masing variabel telah memiliki daun atau keputusan.
4. Membuat Rule berdasarkan pohon keputusan

Berikut tabel 4.2 adalah data dari hasil kuesioner siswa-siswi SMK Analis Abdurrah dan diambil variabel-variabel yang mempengaruhi dalam meminati bahasa asing, data yang dihasilkan dari kuesioner yaitu 100 record. Bahasa asing yang dihasilkan dari kuesioner hanya dua bahasa asing yaitu Inggris dan Arab.

Dari data yang dihasilkan yaitu 100 record, data tersebut akan dicari hitungan *entropy* dan *gain* sesuai langkah-langkah dari algoritma C4.5. Setiap variabel akan ditentukan *entropy*, sebelumnya tentukan dulu *entropy* dari keputusan bahasa yaitu *entropy* total dari seluruh data pada bagian keputusan bahasa.

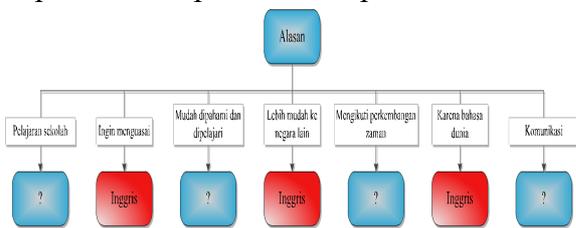
Perhitungan *Entropy*

1. *Entropy* total

Dari data yang dihasilkan yaitu 100 record, data tersebut akan dicari hitungan *entropy* dan *gain* sesuai langkah-langkah dari algoritma C4.5. Setiap variabel akan ditentukan *entropy*, sebelumnya tentukan dulu *entropy* dari keputusan bahasa yaitu *entropy* total dari seluruh data pada bagian keputusan bahasa.

Gain terbesar adalah *Gain* “alasan”. Maka atribut “alasan” menjadi root node atau node akar. Kemudian pada atribut “alasan” ingin menguasai, lebih mudah ke negara lain,

dank arena bahasa dunia memiliki jawaban Inggris. Nilai *Gain* terbesar adalah *Gain* “alasan”. Maka atribut “alasan” menjadi root node atau node akar. Kemudian pada atribut “alasan” ingin menguasai, lebih mudah ke negara lain, dank arena bahasa dunia memiliki jawaban Inggris. Dengan demikian “alasan” ingin menguasai, lebih mudah ke luar negeri, dank arena bahasa dunia menjadi daun atau *leaf*. Maka akan terbentuk pohon keputusan 1 seperti terlihat pada Gambar 5.



Gambar 5. Pohon Keputusan

Berdasarkan Gambar 5 dapat dilihat bahasa Inggris ada pada alasan ingin menguasai, lebih mudah ke negara lain, dan karena bahasa dunia. Berdasarkan pembentukan pohon keputusan 1 (root node), node-node atribut lainnya akan dianalisis lebih lanjut. Dengan memulai menganalisis pada atribut “alasan” mudah dipahami dan dipelajari, mengikuti perkembangan zaman, pelajaran sekolah dan komunikasi dengan melakukan proses yang sama seperti sebelumnya dengan mencari nilai Entropy dan Gain-nya.

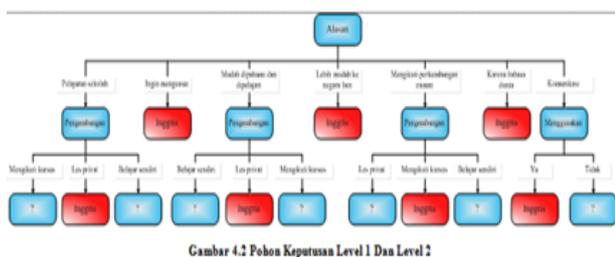
Setelah pada hitungan *entropy* dan *gain* pada level 1 untuk menentukan *root* yang teratas. *Decision tree* yang telah ditentukan maka selanjutnya akan menentukan cabang dari alasan dalam menyukai bahasa asing yaitu karena mengikuti perkembangan zaman. Seterusnya akan dilakukan perhitungan *entropy* dan *gain* sampai setiap cabang menemukan daun atau keputusan

Data cabang dari root alasan yaitu cabang mudah dipahami dan dipelajari. Ini

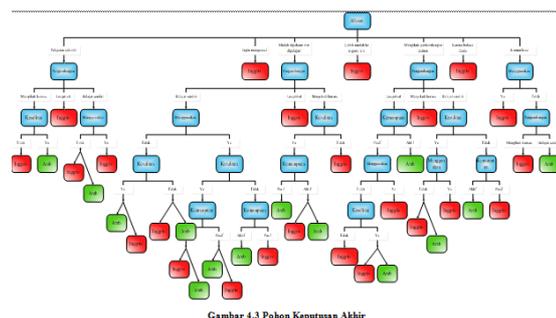
akan dilakukan perhitungan *entropy* dan *gain* sama seperti cabang yang lainnya atau saat akan menentukan *root* pada pohon keputusan dan akan terus dilakukan sampai menemukan keputusan.

Data cabang dari *root* alasan yang ditentukan *entropy* dan *gain* dan begitu seterusnya. Data salah cabang dari *root* alasan yang akan ditemukan hasil keputusan dengan *entropy* dan *gain*.

Gambar 4.2 merupakan pohon keputusan hasil dari perhitungan *entropy* dan *gain* untuk menemukan cabang sampai ke hasil keputusan. Pohon keputusan di atas yang dihasilkan pada level 1 dan level 2. Pohon keputusan belum menemukan daun untuk keputusan akhir. Dan pohon keputusan pada gambar 6 merupakan pohon keputusan



Gambar 4.2 Pohon Keputusan Level 1 Dan Level 2



Gambar 4.3 Pohon Keputusan Akhir

akhir dan telah menemukan daun keputusan. Dalam menemukan daun

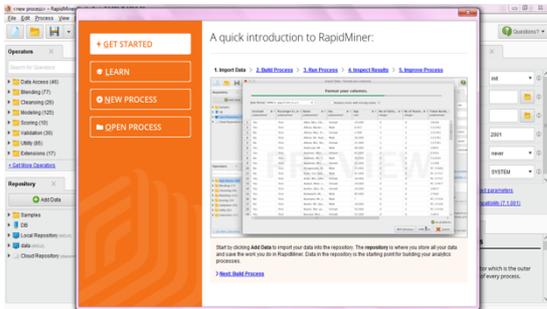
Gambar 6. Pohon Keputusan

keputusan dapat dilakukan perhitungan *entropy* dan *gain* sama hal dalam menemukan *root* dari pohon keputusan.

Implementasi Decision Tree

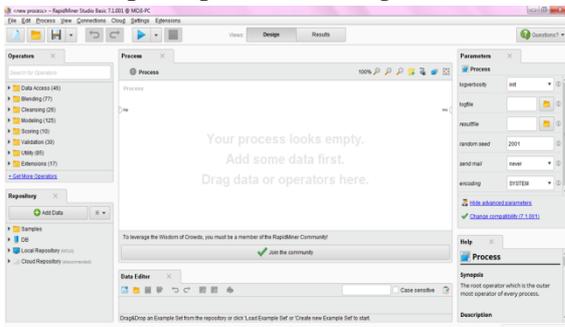
Pengujian terhadap analisa, sangat penting dilakukan untuk menentukan dan memastikan apakah hasil analisa tersebut telah sesuai dengan keputusan yang diharapkan. Untuk menguji kebenaran dari hasil pengolahan data yang dilakukan secara manual, maka dapat menggunakan salah satu software aplikasi *Rapidminer Studio Basic*.

1. Jalankan aplikasi *Rapidminer*
2. Tampilan awal dari *Rapidminer*



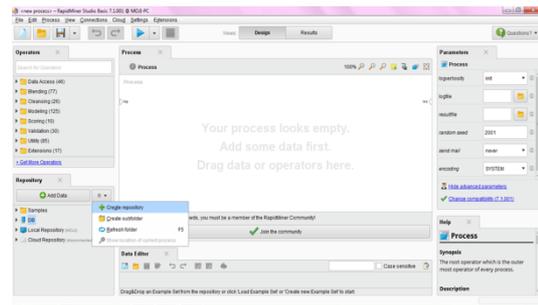
Gambar 7. Tampilan *Rapidminer*

3. Pilihlah *new proses*
4. Tampilan proses dari *Rapidminer*



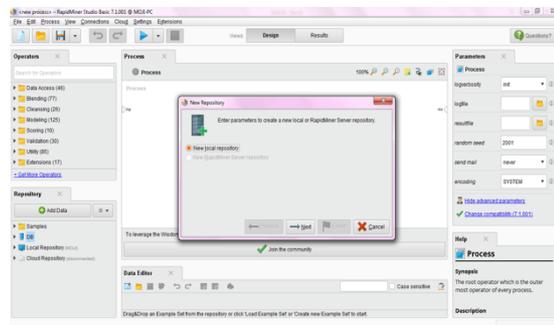
Gambar 8. Tampilan Proses

5. *Create Repository*, tempat penyimpanan data excel yang akan diuji di *rapidminer*



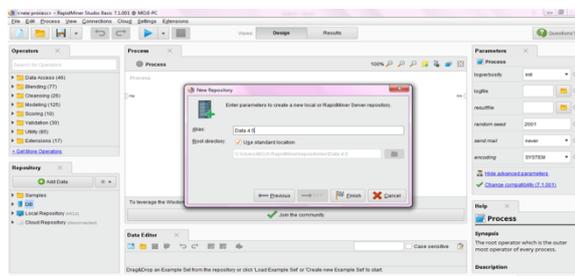
Gambar 9. *Create Repository*

6. Klik Next



Gambar 10. *Create Repository*

7. Ketikkan nama penyimpanan untuk data



uji excel dan klik Finish

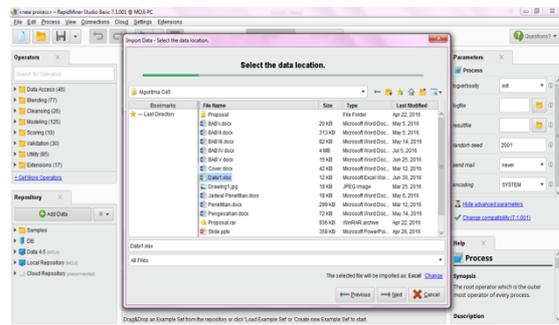
Gambar 11. *Create Repository*

8. Impor data excel yang akan diuji dengan *rapidminer* dengan klik Add Data. Pilih lokasi data excel dengan klik my computer



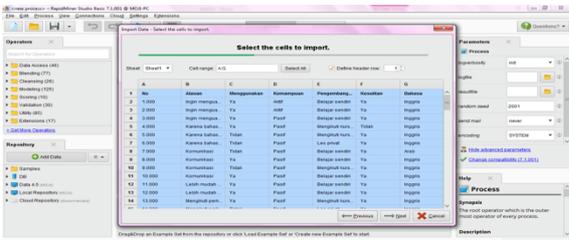
Gambar 12. Import Data Excel

- Pilihlah file Excel yang akan di import ke Rapidminer, dan klik Next



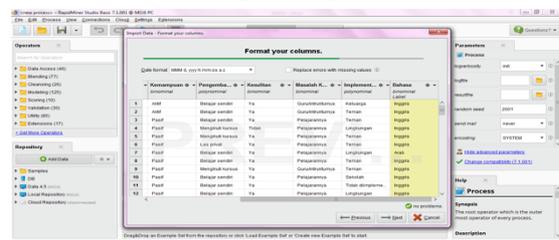
Gambar 13. Import Data Excel

- Pilihlah attribute yang sesuai, klik Next



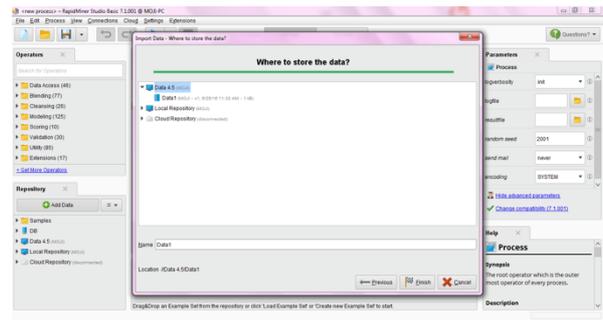
Gambar 14. Import Data Excel

- Rubahlah type yang sesuai, klik Next



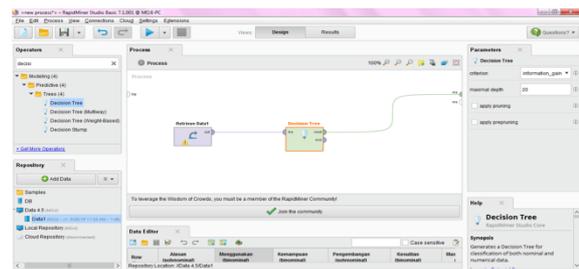
Gambar 15. Import Data Excel

- Beri nama file data Excel, klik Finish



Gambar 16. Import Data Excel

- Drag data excel dan decision tree ke halaman proses. Sambungkan out data Retrieve Data 1 ke tra Decision Tree, dan dari mod decision tree ke res. Pada parameter decision tree bagian criterion pilihlah information_gain dan apply pruning serta apply prepruning dihilangkan tanda centangnya. Setelah klik Run.



Gambar 17. Proses Pohon Keputusan di Rapidminer

KESIMPULAN

Berdasarkan hasil penelitian yang penulis lakukan pada SMK Analis Kesehatan Abdurrah, maka penulis dapat menarik kesimpulan bahwa peminat bahasa asing dengan menggunakan metode Data Mining khususnya Algoritma C4.5 akan bermanfaat sekali dalam proses pengambilan keputusan dalam mengetahui peminat bahasa asing pada siswa-siswi SMK Analis Kesehatan Abdurrah

- Dalam memprediksi peminat bahasa asing menggunakan metode algoritma C4.5 dengan melakukan perhitungan entropy

dan gain dari variabel hasil kuesioner yang dijawab siswa.

2. Variabel yang diperlukan dalam memprediksi peminat bahasa dihasilkan dari hasil survey ke siswa dengan memberikan kuesioner yang akan dijawab. Variabel tersebut adalah alasan meminati bahasa asing, penggunaan bahasa asing, kemampuan bahasa asing, cara mengembangkan bahasa asing, dan kesulitan dalam belajar bahasa asing
3. Menerapkan algoritma C4.5 dalam memprediksi peminat bahasa asing dengan hasil kuesioner yang dijawab siswa berisikan variabel dan hasil jawaban siswa dengan melakukan perhitungan entropy dan gain dalam menentukan faktor tertinggi yang mempengaruhi minat bahasa asing. Dalam memperkuat hasil yang ditemukan dengan menggunakan software dengan mengimport data yang telah disalin atau direkap dari hasil kuesioner. Software yang digunakan adalah RapidMiner. Dengan Rapidminer akan tentukan hasil keputusan. Disana akan tertera hasil dari data yang diimport berupa pohon keputusan.

DAFTAR PUSTAKA

- [1]. Andriani, Anik. (2013). Sistem Pendukung Keputusan Berbasis Decesion Tree Dalam Pemberian Beasiswa (Studi Kasus: AMIK “BSI Yogyakarta”). Seminar Nasional Teknologi Informasi dan Komunikasi 2013 (SENTIKA 2013), Yogyakarta, 9 Maret 2013, ISSN: 2089-9815. AMIK BSI Jakarta
- [2]. Br Ginting, Selvia Lorena., Zarman, Wendi., Hamidah, Ida. (2014). Analisis Dan Penerapan Algoritma C4.5 Dalam Data Mining Untuk Memprediksi Masa Studi Mahasiswa Berdasarkan Data Nilai Akademik. Prosiding Seminar Nasional Aplikasi Sains & Teknologi (SNAST) 2014, ISSN: 1979-911X, 15 November 2014. Yogyakarta.
- [3]. Dosenpendidikan. (2015). 161 Pengertian Bahasa Menurut Para Ahli. (<http://www.dosenpendidikan.com/50-pengertian-bahasa-menurut-para-ahli/>), tanggal akses: 27 Maret 2016.
- [4]. Jefri, & Kusri. (2013). Implementasi Algoritma C4.5 Dalam Aplikasi Untuk Memprediksi Jumlah Mahasiswa Yang Mengulang Mata Kuliah Di STMIK AMIKOM Yogyakarta. *NASKAH PUBLIKASI*. AMIKOM Yogyakarta.
- [5]. Julianto, Windy., Yunitarini, Rika., & Sophan, Mochammad Kautsar. (2014). Algoritma C4.5 Untuk Penilaian Kinerja Karyawan. *SCAN*, Vol: IX, No: 2, ISSN: 1978-0087. Universitas Trunojoyo Madura.
- [6]. Kamagi, David Hartanto., & Hansun, Seng. (2014). Implementasi Data Mining dengan Algoritma C4.5 untuk Memprediksi Tingkat Kelulusan Mahasiswa. *ULTMATICS*, Vol: VI, No: 1, ISSN: 2085-4552. Universitas Multimedia Nusantara Tangerang.
- [7]. Kohar, Abdul. (2011). Minat Siswa Terhadap Kagiatan Ekstrakurikuler Seni Baca Al Aqur’an di MTS Islamiyah Sawangan Depok. Skripsi.
- [8]. Listiana, Mila., Sudjalwa, Gunawan,

- Dedi. (2015). Perbandingan Algoritma Decision Tree (C4.5) Dan Naïve Bayes Pada Data Mining Untuk Identifikasi Tumbuh Kembang Anak Balita (Studi Kasus: Puskesmas Kartasura). Naskah Publikasi. Universitas Muhammadiyah Surakarta. Juli 2015.
- [9]. Prasetyo, Eko. (2014). Data Mining Mengolah Data Menjadi Informasi Menggunakan Matlab. Penerbit ANDI. Yogyakarta.
- [10]. Rahmayuni, Indri. (2014). Perbandingan Performansi Algoritma C4.5 Dan Cart Dalam Klasifikasi Data Nilai Mahasiswa Prodi Teknik Informatika Politeknik Negeri Padang. Jurnal TEKNOIF, Vol: 2, No: 1, April 2014, ISSN: 2338-2724. Universitas Politeknik Negeri Padang.
- [11]. Shita, Rizky Tahara., & Marliani, Nita. (2013). Aplikasi Data Mining Dengan Metode Classification Berbasis Algoritma C4.5. Seminar Nasional Sistem Informasi Indonesia, 2-4 Desember 201. Universitas Budi Luhur.
- [12].Sunjana. (2010). Aplikasi Mining Data Mahasiswa Dengan Metode Klasifikasi Decision Tree. *Seminar Nasional Aplikasi Teknologi Informasi*, Yogyakarta, 19 Juni 2010, ISSN: 1907-5022. Universitas Widyatama.
- [13].Widodo, Prabowo Pudjo., Handayanto, Rahmadya Trias., & Herlawati. (2013). Penerapan Data Mining Dengan Matlab. Penerbit Rekayasa Sains, Bandung.