

Klasifikasi Jurusan SMK Nizam Al-Mulk menggunakan Algoritma C4.5

Alfannisa Annurrallah Fajrin¹, Pastima Simanjuntak², Ellbert Hutabri³
^{1,2,3}Program Studi Teknik Informatika, Universitas Putera Batam, Indonesia

Informasi Artikel

Terbit: Januari 2024

Kata Kunci:

Siswa
Jurusan
Data Mining
Klasifikasi
Algoritma C4.5

ABSTRAK

SMA serta Sekolah Menengah Kejuruan (SMK) merupakan 2 opsi pembelajaran resmi yang berbeda, dimana Sekolah Menengah Kejuruan (SMK) mempunyai beberapa jurusan serta cuma satu jurusan yang bisa dipilih oleh siswa. Sebaliknya SMA lebih sedikit. Beberapa jurusan di Sekolah Menengah Kejuruan (SMK) membuat calon siswa SMP bimbang buatmemilah jurusan yang cocok dengan minatnya. Dalam riset ini penulis hendak membangun sistem pendukung keputusan buat pemilihan jurusan Sekolah Menengah Kejuruan (SMK) bersumber pada data siswa. Informasi Data Mining sendiri mempunyai sebagian metode salah satunya klasifikasi. Metode klasifikasi terdiri sebagian tata cara, serta decision tree merupakan bagian dari tata cara klasifikasi. Setelah itu tata cara decision tree mempunyai algoritma, algoritma C4. 5 merupakan salah satu dari algoritma yang mempunyai decision tree

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Alfannisa Annurrallah Fajrin,
Email: asykhari1302@gmail.com

1. PENDAHULUAN

Semua orang senantiasa dihadapkan dengan sebagian pilihan dalam hidupnya. Apapun keputusan yang diambil hendak mempengaruhi pada kehidupan yang hendak tiba. Begitu pula yang dirasakan oleh siswa SMP (Sekolah Menengah Awal) yang mau melanjutkan pendidikan ke jenjang yang lebih besar. Diantara pilihannya merupakan SMA (Sekolah Menengah Atas) serta Sekolah Menengah Kejuruan (SMK). Sekolah Menengah Kejuruan (SMK) Nizam Al- Mulk tiap tahunnya terus berupaya tingkatkan kualitas pendidikannya dengan harapan lulus bisa mempunyai keahlian lebih dibanding sekolah Sekolah Menengah Kejuruan (SMK) yang lain. Perihal tersebut dicoba buat tingkatkan mutu lulusan Sekolah Menengah Kejuruan (SMK) sehingga siap bersaing didalam dunia kerja. Bersemangat siswa lulusan SMP buat masukke Sekolah Menengah Kejuruan(SMK) lumayan besar, namun banyak siswa kurang matang buat memilah jurusan [1] [2] yang cocok dengan kemampuannya, dampaknya banyak siswa yang kandas [3] ditengah jalur kala mereka telah diterima di sekolah Sekolah Menengah Kejuruan(SMK) tersebut, dan banyak pula permasalahan siswa yang tidak sesuai [4] dengan jurusan yang dipilihnya kala sudah mendapatkan pelajaran disekolah. Jumlah jurusan yang terdapat di Sekolah Menengah Kejuruan(SMK) terdiri dari 3 jurusan ialah jurusan Bahasa, IPA dan IPS. Pemilihan jurusan tiap-tiap jurusan didasarkan pada syarat nilai yang telah ditentukan. Pada tiap jurusan mempunyai kriteria nilai yang sama. Buat jurusan IPA lebih diutamakan nilai

ilmu pasti, semacam matematika dengan nilai rata-rata 75, Fisika, Kimia juga memiliki rata-rata 75. Bersumber pada permasalahan di atas lewat pelaksanaan teknologi informasi mining dengan algoritma C4.5 hingga yang terujimempunyai kinerja yang sangat efisien terhadap data- data yang nilainya berbentuk kisaran dalam jangkauan nilai tertentu.

Informasi Mining ialah penambangan ataupun temuan data baru dengan mencari pola ataupun ketentuan tertentu dari beberapa informasi dalam jumlah besar yang diharapkan bisa menanggulangi keadaan tersebut. Informasi Mining sendiri mempunyai sebagian metode salah satunya klasifikasi[5]. Metodeklasifikasi terdiri sebagian tata cara, serta decision tree merupakan bagian dari tata cara klasifikasi. Setelah itu tata cara decision tree mempunyai algoritma, algoritma C4.5 merupakan salah satu dari algoritma yang mempunyai decision tree [9]. Pohon keputusan merupakan tata cara klasifikasi serta prediksi yang telah teruji powerfull dan sangat populer. Tata cara ini berperan buat mengganti kenyataan jadi tumbuhan keputusan yang merepresentasikan ketentuan yang bisa gampang dipahamidengan bahasa natural.[10]

2. METODE PENELITIAN

2.1 Desain Penelitian

Dalam penentuan jurusan pada SMK Nizam ini ada beberapa hal yang harus dilakukan sebelumnya. Ada beberapa desain penelitian yang merupakan langkah-langkah yang akan dilakukan dalam penyelesaian masalah yang akan dibahas, yaitu :

1. Mendeskripsikan masalah yang akan teliti perlu ditentukan terlebih dahulu. Mendeskripsikan masalah dalam penelitian dengan menentukan dan mendefinisikan batasan masalah yang akan diteliti, sehingga membantu dalam mendapatkan suatu solusi yang terbaik dari masalah tersebut. Jadi, langkah pertama ini adalah langkah awal yang terpenting dalam penelitian ini.
2. Langkah analisis masalah merupakan langkah untuk dapat memahami masalah yang telah ditentukan ruang lingkup atau batasannya. Dengan menganalisis masalah yang telah ditentukan tersebut, maka diharapkan masalah dapat dipahamidengan baik
3. Untuk mencapai tujuan, maka dipelajari beberapa literatur-literatur yang diperkirakan dapat digunakan. Kemudian literatur-literatur yang dipelajari tersebut diseleksi untuk dapat ditentukan literatur-literatur mana yang akan digunakan dalam penelitian.
4. Dalam pengumpulan data dilakukan observasi yaitu pengamatan secara langsung di minimarket ayu sehingga permasalahan yang ada dapat diketahui secara jelas. Kemudian dilakukan *interview* yang bertujuan untuk mendapatkan informasi atau data yang dibutuhkan. Selain itu juga dilakukan studi kepustakaan yaitu dengan membaca buku-buku yang menunjang dalam melakukan analisis terhadap data dan informasi yang didapat. Analisa teknik pengolahan data menggunakan algoritma C4.5 Data yang diperoleh dari tempat penelitian selanjutnya dilakukan analisa dan pengolahan menggunakan algoritma C4.5.
5. Pada tahap ini akan dilakukan proses perancangan dari model sistem dengan algoritma C4.5 sehingga membentuk pohon keputusan (*decision tree*) dan menghasilkan suatu *rule* prediksi Profit Perusahaan
6. Menentukan atribut sebagai akar dan menghitung nilai informasi *gain* atribut, Menyusun *Tree* awal dan Mengubah *Tree* menjadi *rule*.
7. Melakukan pengujian dan hasil perancangan sistem menggunakan *software Data Mining open source RapidMiner*. Sistem diuji dengan prosedur-prosedur untuk melakukan eksplorasi dan permodelan dari data-datayang ada sehingga mendapatkan suatu hubungan tersembunyi dari data tersebut.

2.2 Variabel Penelitian

Variabel yang digunakan pada penelitian dibedakan menjadi variabel bebas dan variabel terikat[11]. Dari datayang dikumpulin, data yang diambil sebagai atribut atau variabel terikat adalah atribut kelayakan anggota security. Atribut-atribut yang digunakan sebagai variabel bebas ada beberapa diantaranya seperti atribut nama, usia, nilai tengah semester, nilai akhir semester, dan sertifikat pelatihan. Penjelasan atribut-atribut yang berperan sebagai variabel bebas yaitu:

1. Nama, berfungsi sebagai variabel ID dengan tipe *text* yang membedakan variabel yang satu dengan yang lain.
2. Nilai UN, sebagai variabel yang menjadi *value* dalam pengolahan sistem penentuan sisa-siswi SMK NAM Batam.
3. Nilai Tes, sebagai bagian variabel dengan *value* perhitungan yang digunakan dalam pengolahan sistem.
4. Tes IQ, berfungsi menjadi variabel dalam penentuan jurusan SMK NAM Batam.
5. Jurusan, menjadi salah satu variabel pilihan awal yang diambil siswa-siswi SMK NAM Batam, yang nantinya setelah penolahan dilakukan menghasilkan keputusan layak atau tidak layak.

Beberapapengembangan yang dicoba pada C4. 5adalahselaku antara lain dapat menanggulangi missing value, dapat menanggulangi continu informasi, serta pruning [8].

3. HASIL DAN ANALISIS

Data-data yang dikumpulkan dari Sekolah Niam AL-Mulk Batam sebanyak 187 data siswa-siswi yang merupakan populasi penelitian. Sampel adalah bagian populasi yang hendak diteliti dan mewakili karakteristik populasi.

3.1. Perhitungan Sample

Apabila populasi penelitian berjumlah kurang dari 100 maka sampel yang diambil adalah semuanya, namun apabila populasi penelitian berjumlah lebih dari 100 maka sampel dapat diambil antara 10-15% atau 20-25% atau lebih. Rumus perhitungan jumlah sampel yaitu:

$$\text{Jumlah sampel} = 20\% \times \text{Jumlah populasi}$$

Rumus 3.1 Perhitungan Sampel

Untuk perhitungan total data populasi sebanyak 100 data dan data yang akan diambil sebagai sampel adalah 20% dari populasi maka perhitungannya adalah sebagai berikut:

$$\text{Jumlah sampel} = 20\% \times 187 = 37,4 \approx 38$$

Pengambilan sampel dalam penelitian ini dilakukan dengan metode *stratified random sampling*. Cara pengambilan sampel secara *stratified random* [6]. Perhitungan kategori sampel yang diperoleh dari rumus 3.2 dapat dilihat pada tabel 3.1 yaitu:

Tabel 3.1 Perhitungan Kategori Sampel

No	Klasifikasi	Perhitungan	Jumlah Sampel
1	IPA	$\frac{38}{187} = 13,4$	13
2	IPS	$\frac{38}{187} = 16,8$	17
3	Bahasa	$\frac{38}{187} = 7,7$	8
Total			38

3.2. Perhitungan Algoritma C4.5

Perhitungan algoritma C4.5 dalam menentukan kelayakan siswa siswi SMK Nizam Al-Mulk berada dalam kelas dengan jurusan yang dipilih dengan cara mencari nilai entropi, *gain*, *split info*, dan *gain ratio*. Pencarian dilakukan secara berulang hingga diperoleh simpul terakhir berupa simpul daun yang tidak bisa diturunkan lagi.

Tabel 3.2 Sample

NO	NAMA	NILAI UN	NILAI TES	IQ	JUR	HASIL
1	Ahmad	85	T	S	0	L
2	Adi	90	T	T	2	L
3	Anto	75	S	R	2	TL
4	Egi	70	R	S	2	L
5	Irwansyah	80	T	T	1	L
6	Supono	90	S	T	2	L
7	Brandon	90	T	T	1	L
8	Ahmad Sobirin	85	S	T	1	L
9	Alek	75	T	S	2	L
10	Putra	70	R	R	2	TL

Dari data pada tabel menghasilkan pengelompokan data tidak layak sebanyak 10 data dan layak sebanyak 28 data, dengan jumlah keseluruhan data sebesar 38 data.

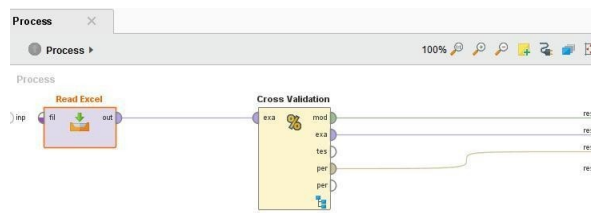
Entropi (total) = 0,8314

3.2. Pembahasan

Pengujian terhadap hasil analisa penting dilakukan untuk memastikan apakah hasil analisa yang dilakukan secara manual benar atau tidak. Untuk menguji kebenaran dari hasil pengolahan data yang dikerjakan maka digunakan salah satu *software* aplikasi dari *data mining* yaitu *RapidMiner*.

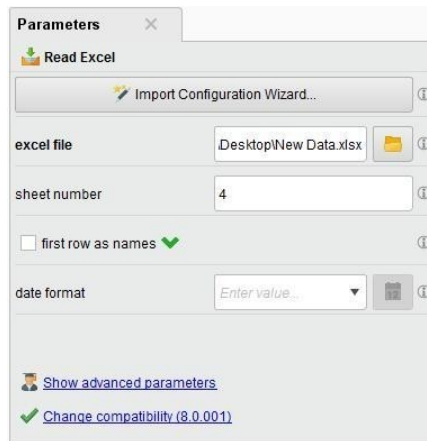
Langkah-langkah implementasi *software RapidMiner* adalah sebagai berikut:

1. Membuka aplikasi *RapidMiner* dan memulai dengan memilih *blank template* maka akan ditampilkan suatu lembar kerja *process* yang kosong.
2. Dalam lembar kerja *process* tersebut, dibuat gambaran sistematis hubungan proses yang akan dilakukan. Dalam penelitian ini, proses yang akan dilakukan yaitu memasukkan data penelitian dan melakukan validasi.



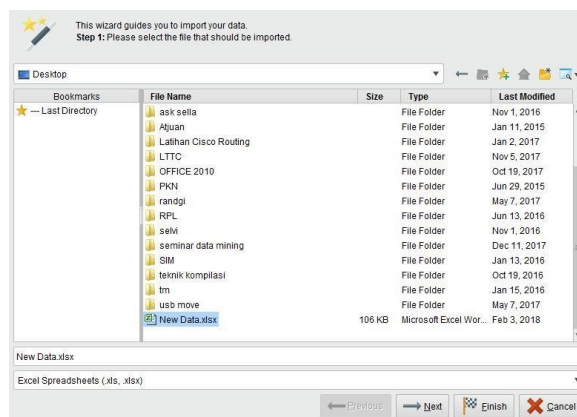
Gambar 3.1 Proses Rapid Miner

3. Selanjutnya klik pada kotak *read excel* dan akan ditampilkan parameter disamping halaman *process*. Halaman parameter akan terlihat pada gambar



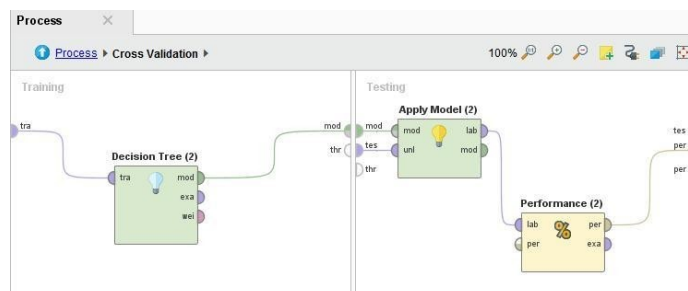
Gambar 3.2 Parameter Data Excel

4. Dalam parameter *read excel* klik pada tombol *import configuration wizard*. Fitur ini digunakan untuk pengambilan data sampel yang sudah diolah untuk diproses *RapidMiner* dari komputer.



Gambar 3.3 Import Data ke *RapidMiner*

5. Selanjutnya gambar adalah gambaran dari isi perancangan kotak *validation*. Disini dibagi menjadi bagian *training* dan bagian *testing*. Pembentukan disesuaikan pada gambar yaitu:



Gambar 3.4 Proses Validasi

Pengujian ini membagi total 38 data sampel menjadi 2 bagian dengan persentase relatif dimana 70% dipakai sebagai data layak dan 30% sisanya sebagai data yang tidak layak.

4. KESIMPULAN

Dari uraian dan penjelasan dari bab-bab sebelumnya, maka dapat disimpulkan bahwa penerapan *data mining* dengan menggunakan algoritma C4.5 cocok digunakan untuk menentukan kelayakan alam penentuan jurusan di SML NAM Batam dengan variabel yang mempengaruhi hasil kelayakannya seperti variabel Nilai UN, Nilai Tes Masuk, IQ, dan Jurusan yang akan dipilih. Algoritma C4.5 memiliki tingkat akurasi yang tinggi dalam menentukan kelayakan penentuan jurusan di SMK Nizam Al-Mulk dengan nilai akurasi sebesar 90,91% dengan pengujian *split validation*, nilai akurasi sebesar 76,32% untuk pengujian. Pengolahan data dari 187 mendapatkann sampel sebanyak 38 sampel dan mendapatkan hasil pengolahan data yang layak sebanyak 28 data, sedangkan data yang tidak layak sebanyak 10 data.

DAFTAR PUSTAKA

- [1] Eka Sabna & Muhardi. (2016). Penerapan Data Mining Untuk Memprediksi Prestasi Akademik Mahasiswa Berdasarkan Dosen, Motivasi, Kedisiplinan, Ekonomi. *CoreIT*, 2(2), 41–44.
- [2] Harryanto, F. F., & Hansun, S. (2017). Penerapan Algoritma C4.5 untuk Memprediksi Penerimaan Calon Pegawai Baru di PT WISE. *Jatiji*, 3(2), 95–103.
- [3] Haryati, S., Sudarsono, A., & Suryana, E. (2015). Implementasi Data Mining Untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma C4.5 (Studi Kasus: Universitas Dehasen Bengkulu). *Jurnal Media Infotama Vol.*, 11(2), 130–138.
- [4] Gamarra, C., Guerrero, J. M., & Montero, E. (2016). A knowledge discovery in databases approach for industrial microgrid planning. *Renewable and Sustainable Energy Reviews*, 60, 615–630.
- [5] Aranski, A. W., & Handoko, K. (2019, November). Data Mining Clustering Nilai IQ Siswa Sekolah SMAK Yos Sudarso Batam. In *Prosiding Seminar Nasional Ilmu Sosial dan Teknologi (SNISTEK)* (Vol.2, pp. 31-36). n.
- [6] Julianto, W., Yunitarini, R., & Sophan, M. K. (2014). ALGORITMA C4.5 UNTUK PENILAIAN KINERJA KARYAWAN, *IX*, 33–39.
- [7] Handoko, K., & Lesmana, L. S. (2018, October). Data Mining Pada Jumlah Penumpang Menggunakan Metode Clustering. In *Prosiding Seminar Nasional Ilmu Sosial dan Teknologi (SNISTEK)* (No. 1, pp. 97-102).
- [8] Irawan, Hendra, (2020). “Inovasi Pendidikan Sebagai Antisipasi Penyebaran Covid-19”.
- [9] Kim, H. joon, Kim, J., Kim, J., & Lim, P. (2018). Towards perfect text classification with Wikipedia-based semantic Naïve Bayes learning. *Neurocomputing*, 315, 128–134. <https://doi.org/10.1016/j.neucom.2018.07.002>
- [10] Kustiyahningsih, Y., & Rahmanita, E. (2016). Aplikasi Sistem Pendukung Keputusan Menggunakan Algoritma C4.5. untuk Penjurusan SMA. *Universitas Trunojoyo*, 5(2), 101–108.
- [11] Meilani, Dwi, B., & Azmuri, W. (2015). Penentuan Pola Yang Sering Muncul Untuk Penerima Kartu Jaminan Kesehatan Masyarakat. *Seminar Nasional “Inovasi Dalam Desain Dan Teknologi,”* 424–431.