

Penerapan Data Mining Dalam Penentuan Jurusan Siswa Dengan Metode Klasifikasi Algoritma C4.5 Studi Kasus SMAN 1 Leuwisadeng

¹Wendy Asswan Cahyadi, ²Subhan Munafis, ³Yose Eduar Muda

^{1,2,3} Program Studi Informatika, STIKom El Rahma

E-mail: ¹wendyasswancahyadi@stikomelrahma.ac.id, ²subhanmunafis@stikomelrahma.ac.id,
³yose.muda@gmail.com

ABSTRACT

The majoring of students in Senior High Schools (SMA) in the 2013 Curriculum is carried out when new students enter class. Often students experience confusion in choosing subjects, due to students' lack of understanding of the subjects they will take. To overcome this, SMAN 1 Leuwisadeng opened direct consultations with BP/BK teachers to provide direction in choosing majors according to students' interests and abilities. However, the major system at SMAN 1 Leuwisadeng is still done manually. As technology develops, this can be overcome with data grouping techniques obtained from data mining results. Data mining is an activity that includes collecting, using data, and history to find regularities, patterns, or relationships in large data sets. In this research, researchers used the C4.5 algorithm to predict students in determining the direction they would take. The accuracy results obtained in this program were 92.31%. Thus it can be concluded that data mining using the C4.5 Algorithm classification method can speed up the process for all students.

Keywords: SMA, Department, Data Mining, Classification, C4.5 Algorithm.

ABSTRAK

Penjurusan siswa di Sekolah Menengah Atas (SMA) pada Kurikulum 2013 dilakukan pada saat siswa baru masuk kelas X. Penjurusan merupakan upaya strategis dalam memberikan fasilitas untuk menyalurkan minat dan kemampuan yang dianggap paling potensial untuk dikembangkan secara optimal. Seringkali siswa mengalami kebingungan untuk memilih jurusan, karena kurang pemahaman siswa terhadap jurusan yang akan diambil. Untuk mengatasi hal tersebut, SMAN 1 Leuwisadeng membuka konsultasi secara langsung kepada Guru BP/BK untuk memberikan arahan dalam memilih jurusan sesuai dengan minat dan kemampuan siswa. Namun sistem penjurusan di SMAN 1 Leuwisadeng masih dilakukan secara manual. Seiring dengan perkembangan teknologi, hal itu dapat diatasi dengan teknik pengelompokan data didapat dari hasil data mining. Data mining adalah kegiatan yang meliputi pengumpulan, pemakaian data, historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar. Dalam penelitian ini, peneliti menggunakan algoritma C4.5 untuk memprediksi siswa-siswi dalam menentukan pilihan jurusan yang akan diambil. Hasil dari akurasi yang didapat pada program ini adalah sebesar 92,31%. Dengan demikian dapat disimpulkan bahwa data mining dengan metode klasifikasi Algoritma C4.5 dapat mempercepat proses penjurusan siswa.

Kata Kunci: SMA, Jurusan, Data Mining, Klasifikasi, Algoritma C4.5.

PENDAHULUAN

Penjurusan merupakan upaya strategis dalam memberikan fasilitas kepada siswa untuk menyalurkan bakat, minat dan kemampuan yang dianggap paling potensial untuk dikembangkan secara optimal. Sehubungan dengan hal tersebut, maka siswa yang mempunyai kemampuan sains dan ilmu eksakta yang baik biasanya akan memilih jurusan MIPA, dan yang memiliki minat pada sosial dan ekonomi akan memilih jurusan IPS, sedangkan yang gemar berbahasa akan memilih jurusan Bahasa [2]. Dengan demikian, karakteristik suatu ilmu menuntut karakteristik yang sama dari yang mempelajarinya. Siswa yang mempelajari suatu ilmu yang sesuai dengan karakteristik kepribadiannya atau minat terhadap suatu ilmu tertentu akan merasa senang ketika mempelajarinya, serta faktor kepribadian mempengaruhi secara positif prestasi akademik. Oleh karena itu, penjurusan bukan masalah kecerdasan tetapi juga masalah minat dan bakat siswa [3].

Di SMAN 1 Leuwisadeng terdapat 2 jurusan, yaitu MIPA dan IPS. Penentuan jurusan berdasarkan minat siswa, dan wawancara dengan Guru BP. Seringkali siswa mengalami kebingungan untuk memilih jurusan, karena kurang pemahaman siswa terhadap jurusan yang akan diambil. Tidak jarang juga siswa-siswi yang asal-asalan dalam menentukan jurusan yang akan mereka ambil, tidak

mempunyai tujuan setelah mereka lulus sekolah menengah atas, dan lain-lain. Hal tersebut berakibat pada menurunnya prestasi belajar mereka, karena ketidakcocokan dengan pelajaran yang dipelajari di jurusan yang dipilih. Untuk mengatasi masalah tersebut siswa hanya berkonsultasi secara langsung kepada wali kelas, BP atau dengan orangtua masing-masing yang dilakukan secara manual sehingga memakan waktu yang cukup lama.

Tetapi seiring dengan perkembangan teknologi hal itu dapat diatasi dengan teknik pengelompokan data didapat dari hasil *data mining*. *Data mining* sering disebut juga sebagai *Knowledge Discovery in Database* (KDD). KDD adalah kegiatan yang meliputi pengumpulan, pemakaian data, historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar [4]. Metode data mining bermacam-macam disesuaikan dengan kebutuhan yang ada. Untuk kasus ini penulis menggunakan algoritma C4.5 untuk memprediksi siswa-siswi dalam menentukan pilihan jurusan yang akan mereka ambil, sehingga nantinya dapat menjadi dasar dalam pengambilan keputusan. Algoritma C4.5 merupakan algoritma klasifikasi dengan teknik pohon keputusan yang terkenal dan disukai karena memiliki kelebihan-kelebihan. Kelebihan ini misalnya dapat mengolah data numerik (kontinyu) dan diskret, dapat menangani nilai atribut yang hilang, menghasilkan aturan-aturan yang mudah diinterpretasikan dan tercepat diantara algoritma-algoritma yang lain [5].

Berdasarkan pemaparan singkat tentang beberapa hal diatas yang menjadi landasan untuk dilakukannya penelitian ini maka penulis membuat aplikasi yang diharapkan dapat menampilkan hasil *data mining* dengan algoritma C4.5 yang berupa pohon keputusan untuk membantu dalam menentukan penjurusan di SMA Negeri 1 Leuwisadeng.

METODE

Desain Penelitian

Desain penelitian adalah suatu strategi untuk mencapai tujuan penelitian yang telah ditetapkan dan berperan sebagai pedoman atau penuntun peneliti pada seluruh proses penelitian. Desain penelitian bagaikan sebuah peta jalan bagi peneliti yang menuntun serta menentukan arah berlangsungnya proses penelitian secara benar dan tepat sesuai dengan tujuan yang telah ditetapkan, tanpa desain yang benar seorang peneliti tidak akan dapat melakukan penelitian dengan baik karena yang bersangkutan tidak mempunyai pedoman arah yang jelas.

Metode Penelitian

Penelitian ini menggunakan metode penelitian kuantitatif yang mencakup studi pustaka, kemudian dilakukan observasi dan pengumpulan data, selanjutnya analisa kebutuhan, kemudian dilakukan perancangan dan terakhir dilakukan pengujian sistem.

Metode Pemilihan Sampel

Populasi yang digunakan oleh penulis adalah seluruh peserta didik baru dari SMAN 1 Leuwisadeng yang belum memilih penjurusan kelas sebelum masuk ke kelas X. Sampel yang digunakan oleh penulis adalah 252 peserta didik baru dengan data minat dan nilai rata-rata raport SMP pada mata pelajaran Matematika, IPA dan IPS semester I sampai dengan semester V.

Metode Pengumpulan Data

Dalam penelitian ini penulis menggunakan data sekunder dimana data yang akan diolah adalah data telah ada. Untuk mendapatkan data tersebut peneliti menggunakan beberapa metode pengumpulan data antara lain:

1. Studi Pustaka

Merupakan metode pengumpulan data yang diperoleh dari buku, jurnal, studi literatur maupun referensi lain yang membahas tentang penjurusan siswa dengan menggunakan Algoritma C4.5.

2. Observasi
Merupakan metode pengumpulan data yang diperoleh dari pengamatan secara langsung di lokasi penelitian untuk mengetahui kondisi yang terjadi.
3. Wawancara
Merupakan teknik pengumpulan data dengan cara berkomunikasi langsung dengan guru Bimbingan Konseling yang bertugas memberi arahan kepada siswa dalam memilih jurusan.

HASIL DAN PEMBAHASAN

Pengumpulan Data

Pada penelitian ini dilakukan proses pengumpulan data dan penyeleksian data dimana data yang sudah didapatkan dari Guru Bimbingan Konseling SMAN 1 Leuwisadeng, Ibu Riana Damayanti, S.Pd., akan dilakukan proses validasi untuk menemukan dan mengkonversi data agar dapat digunakan dalam algoritma data mining dan memperoleh akurasi serta performa yang baik adapun proses validasi yang digunakan dalam algoritma *data mining* dan memperoleh akurasi serta performa yang baik adapun proses validasi yang digunakan tahap ini yaitu menghapus data yang tidak lengkap atau yang tidak memiliki nilai.

Setelah melakukan proses penghapusan data yang tidak memiliki nilai, kemudian langkah selanjutnya adalah penyeleksian atribut, langkah tersebut merupakan proses pemilihan atribut yang dibutuhkan dalam penentuan jurusan siswa, setelah proses pemilihan atribut maka atribut yang akan digunakan yaitu: *Minat, nilai rata-rata Matematika, nilai rata-rata IPA, dan nilai rata-rata IPS*. Sebelum digunakan dan diproses maka data-data tersebut akan disiapkan (*preprocessing*) terlebih dahulu, sehingga nantinya akan memiliki kelas *binominal* atau *polynomial* sesuai aturan yang telah dibuat dan berdasarkan nilai datanya. Adapun data hasil dari *preprocessing* sesuai dengan jenis kelas datanya dapat dilihat pada Tabel 1.

Tabel 1. Potongan Hasil *Preprocessing* (25 record)

No	Nama	Peminatan	Nilai Rata-Rata Raport		
			MTK	IPA	IPS
1	Aat Parihat	MIPA	83,2	86,02	84
2	Abdullah Hamid Zharif	MIPA	79,4	79,4	81,8
3	Ade Saputra	MIPA	85,72	84,13	84,2
4	Adi Saprudin	IPS	75,2	77,2	84,4
5	Afrizal Apriandi Mulya	IPS	85,2	86,4	84,2
6	Agus Miftah	IPS	79	79,43	78,8
7	Ahmad Bustomi	IPS	76,8	76,2	76,4
8	Ahmad Dzakwan Azis	MIPA	80,2	80,4	79,89
9	Ahmad Fauzan Hamdi	MIPA	87,2	87,8	83,2
10	Ahmad Haetami	IPS	87,8	87,4	87,6
11	Ahmad Saripudin	IPS	84	83	84,2
12	Ajril Septian	IPS	82	81,76	80,2
13	Albar Aurelly Setiawan	IPS	72,6	73,8	76,8
14	Alfalah	IPS	80,8	80	81,6
15	Alhaq Sabillilizzati	IPS	80,4	79,8	80,8
16	Alifa Najma Maulida	IPS	85,48	82,1	82,4
17	Alpi Siti Nurrohmah	IPS	77,8	79,87	78,67
18	Alya Ammara	MIPA	89,2	88,4	87,6
19	Alya Suherti	IPS	82,2	88	85,8
20	Amelia Giska Puspita	MIPA	72	75,6	75
21	Amelia Natasya	MIPA	85,2	89	85,6
22	Andika	IPS	81	82,6	80,2
23	Andrian Syahputra	IPS	81,4	87	86
24	Anggi Pratama	IPS	87,4	88,8	83,58
25	Anggini Larasati	IPS	77,6	80,2	77,6

Klasifikasi Data Dengan Algoritma C4.5

Pada tahap penelitian ini dilakukan perhitungan manual menggunakan C4.5 dengan 252 data, adapun langkah-langkah klasifikasi data tersebut dengan menggunakan algoritma C4.5 sebagai berikut:

- Data yang digunakan adalah 252 data, dibagi menjadi 200 data training dan 52 data testing.
- Setelah data disiapkan, maka dilakukan proses perhitungan jumlah siswa yang meminati jurusan MIPA dan IPS berdasarkan nilai atribut.
- Setelah melakukan proses perhitungan tersebut, maka dilakukan perhitungan nilai entropy total pada data training dimana diketahui pada proses perhitungan sebelumnya jumlah siswa yang jurusan MIPA sebanyak 122 siswa dan yang jurusan IPS sebanyak 78 siswa. Perhitungan nilai entropy total tersebut adalah sebagai berikut $Entropy(S) = \sum_{i=1}^n -p_i * \log_2 p_i = (-122/200) * \log_2(122/200) + (-78/200) * \log_2(78/200) = 0.965$.
- Hitung juga nilai entropy pada masing-masing variabel. Berikut adalah salah satu perhitungan nilai entropy pada variabel minat. Nilai entropy pada variabel minat = MIPA adalah sebagai berikut $Entropy(S) = \sum_{i=1}^n -p_i * \log_2 p_i = (-81/82) * \log_2(81/82) + (-1/82) * \log_2(1/82) = 0,095$ Nilai entropy pada variabel minat = IPS adalah sebagai berikut $Entropy(S) = \sum_{i=1}^n -p_i * \log_2 p_i = (-41/118) * \log_2(41/118) + (-77/118) * \log_2(77/118) = 0,932$.
- Setelah dilakukan perhitungan nilai entropy, maka dilakukan juga perhitungan nilai gain untuk masing-masing variabel sehingga diperoleh nilai gain tertinggi nantinya. Perhitungan nilai gain dilakukan untuk menentukan variabel dengan nilai gain tertinggi guna menjadikan variabel tersebut sebagai akar. Berikut adalah contoh salah satu variabel yang dilakukan perhitungan nilai gain nya. $Gain(S, A) = Entropy(S) - \sum_{i=1}^n |S_i| |S| * Entropy(S_i) = 0,965 - (((82/200)*(0,0950) + ((118/200)*0,932)) = 0,376$.

Untuk memperoleh nilai gain tertinggi dimana nilai gain yang diperoleh akan dijadikan sebagai akar maka harus ditentukan oleh hasil perhitungan nilai entropy dan gain dari semua atribut.

Evaluasi Model dengan Confusion Matrix

Penelitian ini melakukan proses evaluasi model dengan menggunakan confusion matrix, dengan proses tersebut akan terbentuk matrix yang terdiri dari accuracy, precision dan recall. Dengan data training yang sudah disiapkan ke dalam confusion matrix maka pengujian pertama sekali dilakukan dengan memperoleh nilai confusion matrix yang diperoleh dari tools Rapid Miner. Berikut adalah hasil perhitungan confusion matrix untuk mencari nilai accuracy, precision dan recall dari metode C4.5.

Tabel 2. Nilai accuracy, precision, dan recall

	True MIPA	True IPS
Pred MIPA	132	14
Pre IPS	12	94

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} = \frac{132+94}{132+94+12+14}$$

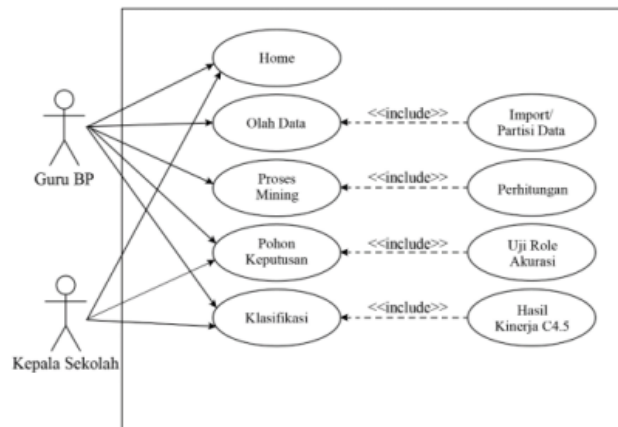
$$= \frac{226}{252} \times 100\% = 89,68\%$$

$$Precision = \frac{TP}{(TP+FP)} = \frac{132}{132+12} = \frac{132}{144} \times 100\% = 91,67\%$$

$$Recall = \frac{TP}{(TP+FN)} = \frac{132}{132+14} = \frac{132}{146} \times 100\% = 90,41\%$$

Perancangan Sistem

Dalam perancangan model, penulis menggunakan 3 model UML, yaitu: *Use Case Diagram*, *Class Diagram* dan *Activity Diagram*. *Use case diagram* yang dibuat untuk menggambarkan kebutuhan fungsional pada penelitian ini dapat dilihat pada Gambar 1.



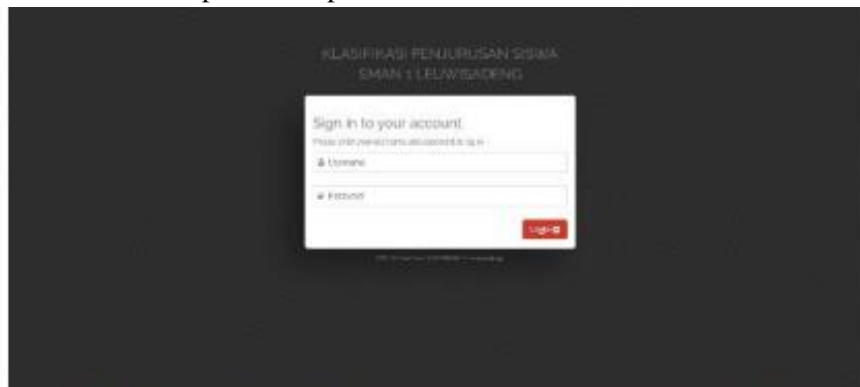
Gambar 1. Use Case Diagram

Implementasi Sistem

Berikut merupakan implementasi dari rancangan sistem yang telah dibuat.

1. Halaman Login

Halaman login diperlukan untuk mengetahui hak akses pengguna sebelum masuk ke dalam sistem. User melakukan input username dan password, jika sesuai dengan data yang tersimpan dalam database, maka akan berhasil masuk ke halaman utama. Tampilan halaman login pada sistem ini dapat dilihat pada Gambar 2.



Gambar 2. Halaman Login

2. Halaman Utama

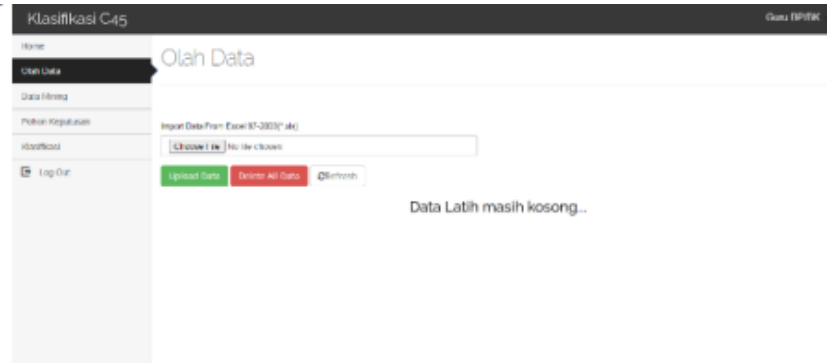
Halaman utama ditampilkan untuk pengguna saat berhasil masuk/login ke sistem. Tampilan halaman utama pada sistem ini dapat dilihat pada Gambar 3.



Gambar 3. Halaman Utama

3. Halaman Olah Data

Halaman olah data ditampilkan untuk pengguna saat mengolah data penjurusan. Tampilan halaman olah data pada sistem ini dapat dilihat pada Gambar 4.



Gambar 4. Halaman Olah Data

4. Halaman Data Mining

Halaman data mining ditampilkan untuk pengguna saat memproses data penjurusan dengan metode data mining klasifikasi algoritma C4.5. Tampilan halaman data mining dan proses mining dapat dilihat pada Gambar 5.

No	Nama	Jurusan	Rata-rata MTK	Rata-rata IPA	Rata-rata IPS	Jurusan
1	AM FARIHAT	MIPA	83.3	86.5	84	MIPA
2	ANDULIAH HAWID ZHARIF	MIPA	79.4	75.1	81.8	MIPA
3	ACE SAPUTRA	MIPA	82.7	84.1	84.2	MIPA
4	ACH KAPRISON	IPS	78.2	77.2	81.4	IPS
5	ARTIYAL AFFANED MUBLA	IPS	85.2	86.4	84.2	MIPA
6	AGUS MITANI	IPS	79	75.4	79.8	IPS
7	AHMAD RAHMAS	IPS	78.8	78.2	76.4	IPS
8	AL-BAND UDHOMAN AQIB	MIPA	82.2	82.4	79.8	MIPA
9	AHMAD FRIYAN HANID	MIPA	87.2	87.8	83.2	MIPA

Gambar 5. Halaman Data Mining

5. Halaman Pohon Keputusan

Halaman pohon keputusan ditampilkan untuk pengguna saat melihat hasil perhitungan berupa pohon keputusan. Dalam halaman pohon keputusan juga ada tombol Uji Rule untuk menguji tingkat akurasi pohon keputusan. Tampilan halaman pohon keputusan dan uji rule dapat dilihat pada Gambar 6.

No	Aturan
1	IF (rata_rata_IPA >= 80) AND (rata_rata_IPS <= 80) THEN Label = MIPA
2	IF (rata_rata_IPA <= 80) AND (rata_rata_IPS <= 80) AND (rata_rata_IPA <= 80) THEN Label = MIPA
3	IF (rata_rata_IPA <= 80) AND (rata_rata_IPS <= 80) AND (rata_rata_IPA <= 80) AND (rata_rata_IPS <= 80) THEN Label = MIPA
4	IF (rata_rata_IPA <= 80) AND (rata_rata_IPS <= 80) AND (rata_rata_IPA <= 80) AND (rata_rata_IPS <= 80) THEN Label = MIPA
5	IF (rata_rata_IPA <= 80) AND (rata_rata_IPS <= 80) AND (rata_rata_IPA <= 80) AND (rata_rata_IPS <= 80) THEN Label = MIPA
6	IF (rata_rata_IPA <= 80) AND (rata_rata_IPS <= 80) AND (rata_rata_IPA <= 80) THEN Label = MIPA
7	IF (rata_rata_IPA <= 80) AND (rata_rata_IPS <= 80) AND (rata_rata_IPA <= 80) THEN Label = IPS
8	IF (rata_rata_IPA <= 80) AND (rata_rata_IPS <= 80) AND (rata_rata_IPA <= 80) THEN Label = IPS
9	IF (rata_rata_IPA <= 80) AND (rata_rata_IPS <= 80) AND (rata_rata_IPA <= 80) THEN Label = IPS
10	IF (rata_rata_IPA <= 80) AND (rata_rata_IPS <= 80) AND (rata_rata_IPA <= 80) THEN Label = MIPA

Gambar 6. Halaman Pohon Keputusan

KESIMPULAN

Dari hasil penelitian dan pembahasan yang dilakukan, didapatkan kesimpulan berikut:

1. Dengan sistem penjurusan siswa ini diharapkan mampu membantu guru BK dan siswa dalam menentukan jurusan siswa sesuai proses perhitungan data mining.
2. Metode klasifikasi C4.5 yang digunakan pada sistem ini memiliki nilai akurasi 92,31%, sehingga metode ini tepat untuk mengklasifikasikan data penjurusan siswa.

DAFTAR PUSTAKA

- Achmad Solichin. 2016. "Pemrograman Web dengan PHP dan MySQL". Hal 10
- Agung Baitul Hikmah, dkk. 2015. Cara Cepat Membangun Website dari Nol Studi Kasus : Web Dealer Motor. Yogyakarta: CV. Andi Offset.
- Ali Zaki dan Smithdev Community. 2008. Belajar Komputer PHP dan MySQL. Semarang: Elex Media Komputindo.
- Basuki, A., dan Syarif, I. 2003. Pohon Keputusan. <http://lecturer.eepis-its.edu/~basuki/lecture/DecisionTree.pdf>
- Beti Novianti, Tedy Rismawan dan Syamsul Bahri. 2016. Implementasi Data Mining dengan Algoritma C4.5 untuk Penjurusan Siswa (Studi Kasus SMA Negeri 1 Pontianak).
- Dhika, Harry. 2015. Kajian Komparasi Penerapan Algoritma C4.5, Naïve Bayes, dan Neural Network dalam Pemilihan Kerja Penyedia Jasa Transportasi: Studi Kasus CV. Viradi Global Pratama. SNIT.
- Eka Budi Rayahu. Algoritma C4.5 untuk Penjurusan Siswa SMA Negeri 3 Pati.
- Han, J. dan M. Kamber. 2006. Data Mining Concepts and Techniques Second Edition. San Francisco: Morgan Kaufmann.
- Kusrini dan Luthfi, E. T. 2009. Algoritma Data Mining. Yogyakarta: Andi Offset.
- Mantas, C. J., & Abellan, J. 2014. Credal-C4.5: Decision tree based on imprecise probabilities to classify noisy data. Expert Systems with Applications, 41(10), 4625-2637.doi:10.1016/j.eswa.2014.01.017
- Pressman, Ph.D. Roger S. 2010. Pendekatan Praktisi Rekayasa Perangkat Lunak Edisi 7. Penerbit Andi. Yogyakarta
- Purbadian, Yenda. 2016. Trik Cepat Membangun Aplikasi Web dengan Framework Codeigniter. Yogyakarta: CV. Andi Offset
- Quinlan, J.R. 1986. Induction of Decision Trees. Machine Learning, 81-106
- Ramli, Murni. 2008. Penjurusan di SMA. <http://indosdm.com/penjurusan-di-sma>
- Rina Kurniasari dan Azizah Fatmawati. 2019. Penerapan Algoritma C4.5 untuk Penjurusan Siswa Sekolah Menengah Atas.
- Santoso, Budi. 2007. Data Mining: Teknik Pemanfaatan data untuk keperluan bisnis. Yogyakarta: Graha Ilmu
- Snow, RE. 1986. Individual Differences And The Design Of Educational Program In Journal Of Psychology. American Psychologist, Vol 41(10), Oct 1986, 1029-1039
- Supono, dan Virdiandry Putratama. 2016. Pemrograman Web dengan Menggunakan PHP dan Framework Codeigniter. Yogyakarta: Deepublish (Grup Penerbitan CV Budi Utama).
- Tang, ZhaoHui and J. MacLennan. 2005. Data Mining With SQL Server 2005. Indianapolis: Wiley Publishing, Inc.
- Turban, E., dkk. 2005. Decision Support Systems and Intelligent Systems. Yogyakarta: Andi Offset.