

KLASIFIKASI DENGAN DECISION TREE UNTUK PENENTUAN MODEL PEMBELAJARAN YANG EFEKTIF STUDI KASUS SMK GAMA CENDIKIA SURABAYA

Sehman*

* Teknik Informatika, Fakultas Teknik
Universitas Maarif Hasyim Latif Sidoarjo

Correspondence Author: sehman@dosen.umaha.ac.id

| Info Artikel : | ABSTRAK |
|---|---|
| <p>Sejarah Artikel : Menerima 16-01- 2019 Revisi 03-02-2018 Diterima 03-02-2019 Online 04-02-2019</p> <p>Key words: <i>Learning, Model, Process, Decision, Teacher.</i></p> | <p><i>The learning model is one of the main components in creating an atmosphere of active learning, innovative, creative and fun. learning model refers to the approach to be used. To that end, established a system of analysis by using the methods in data mining are: ID3 which is expected to provide solutions for the teacher to choose a model for effective learning. The purpose of the research in order to contribute to the vocational Gama Scholar and the teachers and students generally. The final result or output of this research is the electoral system learning model based on inputs such as the circumstances of the learning process in vocational Gama Scholar Surabaya, as for the type of learning process is the specialization of educators teachers, iq students, facilities, time, atmosphere, objectives, and type of material.</i></p> |
| INTISARI | |
| <p>Kata Kunci : <i>Model, Pembelajaran, Proses, Decision, Guru</i></p> | <p><i>Model pembelajaran adalah salah satu komponen utama dalam menciptakan suasana belajar yang aktif, inovatif, kreatif dan menyenangkan. model pembelajaran mengacu pada pendekatan yang akan digunakan. Untuk itu, dibentuklah suatu sistem analisa dengan menggunakan metode- metode pada data mining yaitu: ID3 yang diharapkan dapat memberikan solusi bagi pihak guru untuk memilih model pembelajaran yang efektif. Tujuan dilakukan penelitian guna memberikan kontribusi kepada siswa SMK dan guru- guru serta para siswa umumnya Hasil akhir ataupun output dari penelitian ini adalah system pemilihan model pembelajaran berdasarkan input yang berupa situasi dan kondisi proses pembelajaran di SMK, adapun jenis proses pembelajaran ialah spesialisasi tenaga pendidik/guru, iq siswa, fasilitas, waktu, suasana, tujuan, dan jenis materi.</i></p> |

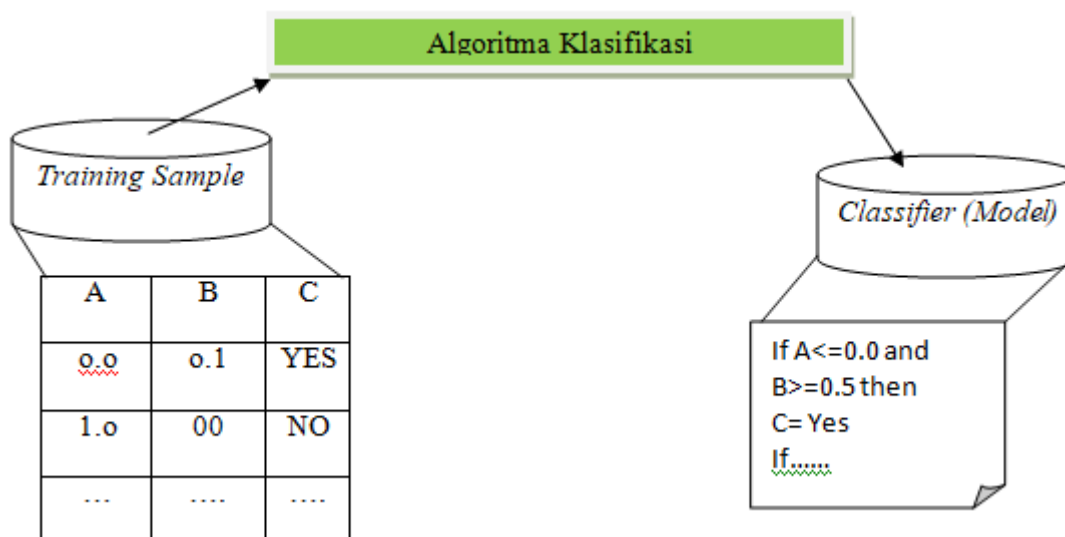
1. PENDAHULUAN

Tercapainya tujuan pembelajaran tentunya dipengaruhi oleh aktifitas siswa dan guru selama proses belajar mengajar. Pemilihan model pembelajaran sangat mempengaruhi terhadap tingkat keberhasilan siswa dalam mencapai tujuan pembelajaran yang telah ditetapkan oleh undang undang dasar. Model pembelajaran merupakan pola atau rencana yang dapat digunakan untuk mengoperasikan kurikulum, merancang materi pembelajaran, dan untuk membimbing belajar dalam setting kelas atau lainnya (Suprijono, 2013). Model pembelajaran ialah pola yang digunakan sebagai pedoman dalam merencanakan pembelajaran dikelas maupun tutorial. model pembelajaran

mengacu pada pendekatan yang akan digunakan, termasuk di dalamnya tujuan-tujuan pembelajaran, tahap dalam kegiatan pembelajaran, lingkungan pembelajaran, dan pengelolaan kelas (Slavin 2010). Banyak model pembelajaran yang bisa diterapkan, akan tetapi perlu mempertimbangkan beberapa faktor seperti: Spesialisasi Tenaga Pendidik/Guru, Kompetensi Dasar, IQ siswa, Fasilitas, Waktu, Suasana, Tujuan, dan Jenis Materi. Selama ini faktor-faktor tersebut tidak diperhatikan dengan maksimal dalam pemilihan model pembelajaran yang akan digunakan pada proses belajar mengajar, sehingga proses pembelajaran di kelas tidak berjalan dengan efektif. Hal ini dibuktikan oleh data rencana perangkat pembelajaran guru SMK Gamam Cendikia Surabaya yang diambil berdasarkan data 4 tahun terakhir yaitu dari tahun pembelajaran 2012/2013 sampai dengan tahun pembelajaran 2015/2016. Melihat realita tersebut, maka sangat dibutuhkan keberadaan suatu sistem yang bisa membantu guru dalam pemilihan model pembelajaran yang efektif.

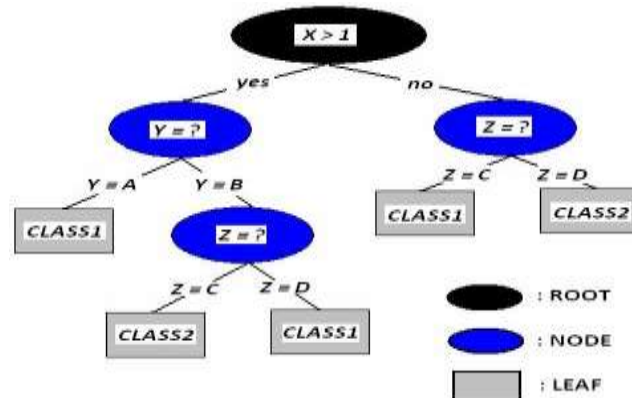
2. METODE PENELITIAN

Pembuatan model, pada tahap ini setiap data diasumsikan telah digolongkan kedalam sejumlah kelas (*predefined class*). Himpunan data yang akan menyusun model ini disebut sebagai training data. Model yang dihasilkan dipresentasikan dalam bentuk aturan klasifikasi, pohon keputusan atau formula matematika.



Gambar 1. Ilustrasi Pembuatan Pohon Keputusan

Representasi pohon keputusan ini dianggap sebagai metode logis yang sering digunakan pada bahasan mengenai statistik terapan dan pembelajaran mesin (*machine learning*). Pembuatan pohon keputusan sendiri menggunakan metode *supervised learning* yaitu proses pembelajaran dimana data ini terdiri dari nodes atau simpul yang merupakan atribut dari data sampel. Cabang (*branches*) yang keluar dari node tersebut merupakan nilai atau *outcome* yang dimiliki oleh atribut (*nodes*) bersangkutan. Sedangkan daun yang ada pada pohon keputusan tersebut adalah kelas dari data sampel yang diuji (Kusrini, 2009). Sebagai ilustrasi dapat dilihat pada contoh gambar berikut.



Gambar 2. Model Pohon Keputusan

Sebelum ekstraksi data kedalam bentuk model *tree* tentunya ada beberapa proses yang harus diperhatikan dalam pembentukan struktur pohon ini, yaitu:

1. Penghitungan Nilai Entropy

Untuk mendapatkan nilai information gain, terlebih dahulu harus mengetahui parameter lain yang mempengaruhi nilai gain, dimana parameter ini sangat diperlukan untuk mendapatkan nilai gain. Parameter tersebut adalah entropy. Parameter ini sering digunakan untuk mengukur heterogenitas suatu kumpulan sampel data. Secara matematis nilai entropy dapat dihitung dengan menggunakan Persamaan 1.

$$E(S) = -p + \log_2(p+) - p - \log_2(p-) \dots\dots\dots(1)$$

Keterangan:

- S = Himpunan kasus atau ruang (data) sampel yang digunakan untuk training.
- P+ = Jumlah yang bersolusi positif (mendukung) pada data sampel untuk kriteria tertentu.
- P- = Jumlah yang bersolusi negative (tidak mendukung) pada data sampel untuk kriteria tertentu.

Entropy (S) adalah jumlah bit yang diperkirakan dibutuhkan untuk dapat mengekstrak suatu kelas (+ atau -) dari sejumlah data acak pada suatu ruang sampel S. Entropy bisa dikatakan sebagai kebutuhan bit untuk menyatakan suatu kelas. Semakin kecil nilai dari suatu entropy maka semakin baik digunakan dalam mengekstraksi suatu kelas. Setelah mendapatkan nilai entropy, maka langkahselanjutnya adalah melakukan perhitungan terhadap information gain. Dengan menggunakan persamaan 2.

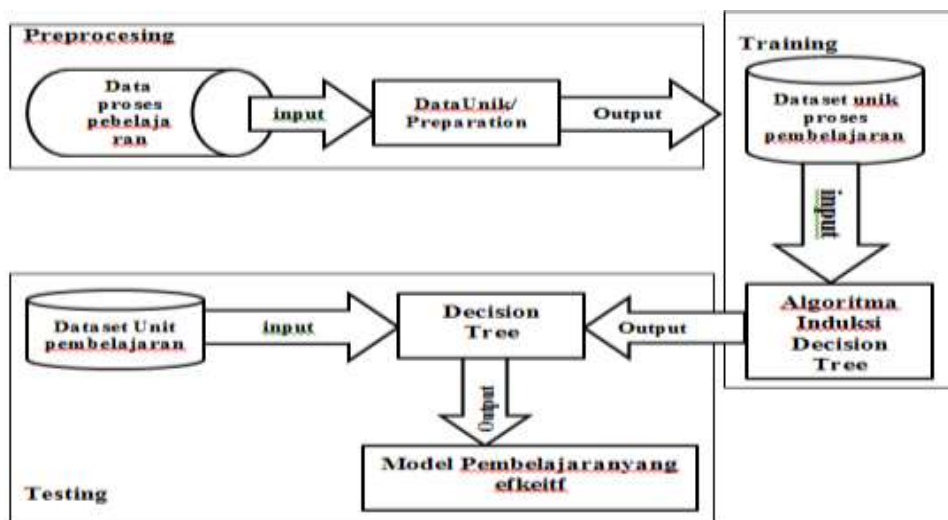
$$G(S, A) = E(S) - \sum_{v \in V(A)} \frac{Sv}{S} E(Sv) \dots\dots (2)$$

Adapun dataset proses pembelajaran SMK Gama Cendikia Surabaya bersasal dari Tata usaha yang berisi data spesialisasi tenaga pendidik/guru, iq siswa, fasilitas, waktu, suasana, tujuan, dan jenis materi. Adapun dataset proses pembelajaran di tunjukkan pada tabel 1. Dengan Jumlah data set keseluruhan ialah 88.080 data dengan 7.340 record dan 12 field. Setelah dataset ditetapkan maka akan dilakukan tahapan pengolahant data yaitu tahapan *preprocessing*. Adapun dataset yang akan *dipreprocessing* ditunjukkan pada tabel 1.

Tabel 1. Dataset

| Variabel Input | Keterangan |
|--------------------|--------------|
| Mata Pelajaran | Input/ |
| Kompetensi Dasar | Input |
| IQ Siswa | Input |
| Tingkat Siswa | Input |
| Fasilitas | Input |
| Kemampuan Guru | Input |
| Jenis Materi | Input |
| Waktu | Input |
| Suasana | Input |
| Tujuan | Input |
| Nilai | Prediksi |
| Model Pembelajaran | Class/target |

Adapun proses pengolahan data ditunjukkan pada gambar 3.



Gambar 3. Proprocessing

Setelah dilakukan *preprocessing* maka data mengalami perubahan menjadi 23.814 dengan *record* 2.646 dan 9 *field*. hasil dari pengolahan data yang telah sukses dibuktikan dengan format arff. setelah dataset sukses diolah maka akan dilakukan pengujian model data mining atau uji coba.

3. HASIL DAN ANALISA

Adapun uji coba yang dilakukan ialah

1. Pembagian Data set

Dataset akan dibagi menjadi 2 subset yaitu data training, dan data testing Dataset original dengan jumlah 23.814 dengan *record* 2.646 dan 9 *field*.

2. Training Set

Dari dataset original tersebut akan diambil 80% untuk *data training* dengan jumlah *instances* 2.116, *data testing* 20% dengan jumlah 529 *instances*. Fungsi dilakukannya training data

ialah untuk membentuk model dan melakukan pelatihan terhadap data tersebut kemudian dari data training akan diuji model dengan menggunakan data testing. Seperti yang ditunjukkan pada gambar 4.

```

Time taken to build model: 0.04 seconds

=== Evaluation on training set ===
=== Summary ===

Correctly Classified Instances      1823      86.1531 %
Incorrectly Classified Instances    293      13.8469 %
Kappa statistic                    0.8478
Mean absolute error                 0.0313
Root mean squared error             0.1251
Relative absolute error             18.9383 %
Root relative squared error         43.5182 %
Total Number of Instances          2116

```

Gambar 4. Proses Uji Coba

Dari gambar 4 tersebut ditunjukkan bahwa dari 80% memiliki jumlah 2116 *instances* namun setelah dilakukan proses *training* dengan menggunakan algoritma id3 maka ditemukan 1.823 klasifikasi kebenaran *instances* dengan akurasi 86.1531% dan kesalahan klasifikasai *instancse* berjumlah 293 *instances* dengan akurasi 13.8469%. Dari hasil proses yang telah dilakukan dapat disimpulkan hasil dari *training test* cukup baik dan kemudian dilakukan percobaan pada tahap data testing.

3. Testing Set

Dari original data model pembelajaran data yang digunakan untuk melakukan *testing* sebesar 20% dengan jumlah *instances* 529 *instances* dan 9 atribut adapun tahapan yang dilakukan untuk testing set ialah : pada awalnya dataset dibagi terlebih dahulu di weka dengan cara menampilkan seluruh dataset model pembelajaran kemudian melakukan filter, resample dan memasukan nilai simple size percent sebanyak 20% kemudian di apply dan disimpan. Setelah tersimpan data testing tersebut dibuka dengan menggunakan aplikasi berbasis notepad++ guna untuk memeriksa hasil resample sebelumnya. *Testing set.arff* dijelaskan hasil dari *resample* telah berhasil dengan ditunjukanya masing masing input dan atribut.

Tree yang dihasilkan pada dataset proses pembelajaran dengan training data 80 % Pada prediksi nilai cukup diketahui bahwa IQ merupakan node awal dengan demikian dapat dipastikan bahwa nilai entropy tertinggi pada prediksi nilai baik ialah IQ dan gain tertinggi ialah fasilitas. dari hasil tree yang ditampilkan akan menghasilkan rule klasifikasi, adapun contoh rule klasifikasi ialah:

If IQ="Rendah" dan Fasilitas ="Lab" maka model ="Jigsaw"

seorang siswa memiliki IQ rendah dan belajar menggunakan fasilitas lab maka model yang efektif untuk digunakan ialah jigsaw, *Rule* di atas merupakan uji coba yang telah dilakukan dengan menggunakan aplikasi yang dibuat oleh penulis, namun penjelasan tersebut hanya menjelaskan beberapa node dan beberapa leaf dari *tree*. Adapun akurasi yang dihasilkan dari *tree* berdasarkan nilai prediksi dapat dilihat pada tabel 2.

Tabel 2. Akurasi

| | Data training | Data testing | Akurasi |
|--------------------------|---------------|--------------|---------|
| Baik | 80% | 20% | 79.46 % |
| Cukup | 80% | 20% | 76.24 % |
| Sangat baik | 80% | 20% | 76.64 % |
| Sangat Baik Baik, cukup, | 80% | 20% | 79.06 % |

Dari tabel 2 tersebut dapat disimpulkan bahwa hasil akurasi rata-rata adalah 78% dan membuktikan cukup baik. *K-Fold/Cross-Validation* berguna untuk menghindari overlapping pada data testing dalam tahapan ini memiliki 2 proses yaitu : membagi data menjadi K-subset yang berukuran sama dan menggunakan setiap k - subset untuk data testing dan sisanya sebagai *data training*. *Cross Validation Method fold = 5* dengan penjelasan berikut :

Data terdiri dari 2.646 *instances/record* dibagi ke dalam 6 blok dengan jumlah *instances* yang sama yaitu: blok : A, B, C, D dan E, masing-masing terdiri dari 20 *instances*. Kualitas kombinasi parameter tertentu diuji dengan cara tahapan:

step 1: *training* memakai A,B,C,D,F *testing* memakai E akurasi a.

step 2: *training* memakai A,B,C,E ,F *testing* memakai D akurasi b

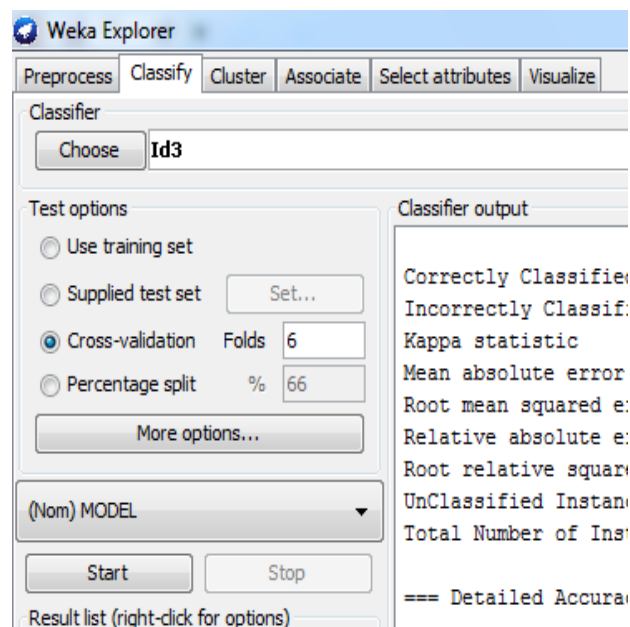
step 3: *training* memakai A,B, D,E,F *testing* memakai C akurasi c

step 4: *training* memakai A, C,D,E ,F *testing* memakai B akurasi d

step 5: *training* memakai B,C,D,E,F *testing* memakai A akurasi e

step 6: *training* memakai B, C, D, E, F *testing* memakai F akurasi f

Rata-rata akurasi: $(a+b+c+d+e+f)/6$ mencerminkan kualitas parameter yang dipilih mengubah parameter model, dan mengulangi dari no.2 sampai dicapai akurasi yang diinginkan.



Gambar 5. Proses algoritma Id3 di weka

Adapun dari proses yang dihasilkan oleh weka ialah *Correctly Classified Instances* 2118 dengan akurasi 80.0454 %, *Incorrectly Classified Instances* 508 dan akurasi 19.1988%, *Kappa statistic* 0.7872, *Mean absolute error* 0.0383, *Root mean squared error* 0.1513, *Relative*

absolute error 23.3653 %, *Root relative squared error* 52.8366 %, *UnClassified Instances* 20 0.7559 %, *Total Number of Instances* 2646, untuk lebih jelas dapat dilihat pada gambar 4.

```

=== Confusion Matrix ===
  a  b  c  d  e  f  g  h  i  j  k  <-- classified as
218  7 11  1  0  0  0  0  0  0  0 | a = Think-Pair-And-Share
  5 159 71  1  0  0  0  0  0  0  0 | b = Ceramah
11  90 138  1  0  0  0  0  0  0  0 | c = CBSA
  2  2  3 231  0  2  0  0  0  0  0 | d = Jigsaw
  0  0  0  0 202  1 36  2  0  0  0 | e = Cooperative
  0  0  0  7  8 193 21  0  0  0  0 | f = Direct Interaction
  0  0  0  0 120 26 106  6  0  0  0 | g = Explicit Intruction
  0  0  0  0  5  0  0 215 21  0  0 | h = Demonstration
  0  0  0  0  0  0  0  10 221 11  0 | i = Inkuiri
  0  0  0  0  0  0  0  0  2 216  5 | j = Numbered Head Togethe
  0  0  0  0  0  0  0  0  0  20 219 | k = FBI

```

Gambar 6. Confusion Matrix

Gambar 4 ialah Baris pertama “218 7 11 1 0 0 0 0 0 0 0” menunjukkan bahwa ada (218 + 7+ 11 + 1+ 0+ 0 + 0 + 0+ 0 + 0+ 0) *instances class* proses pembelajaran di dalam file proses pembelajaran.arff dan semua benar diklasifikasikan sebagai *Think-Pair-And-Share*. Dan demikian pula cara membaca baris baris berikutnya.

4. KESIMPULAN

Jumlah data training dan testing yang digunakan mempengaruhi nilai akurasi pada setiap percobaan. Metode *decision tree* dapat melakukan klasifikasi jika dataset bebas dari gangguan/noise dan tahapan yang membutuhkan waktu lama ialah fase preprocessing. Berdasarkan uji coba 80% untuk *training set* dan 20% untuk *testing set* dengan perhitungan ID3 menghasilkan akurasi 80%. IQ merupakan atribut yang memiliki nilai entropy tertinggi dengan demikian *variable* yang paling berpengaruh terhadap model pembelajaran ialah IQ.

DAFTAR PUSTAKA

- Agus Suprijono. 2013 .*Cooperative Learning Teori dan Aplikasi PAIKEM*. Yogyakarta: Pusaka Pelajar
- Jiawei Han, Micheline Kamber, Jian Pei, *Data Mining-Concepts and Techniques (third dedition)*, New york, Morgan Kaufmann
- Kusrini, Luthfi Taufiq Emha, (2009), *Algoritma Data Mining*, Penerbit Andi, Yogyakarta.
- Michael J.A. Berry, Gordon S.Linoff , *Data Mining Technic For Marketing Sales And Costumer Support*, New York, JohnWiley & Sons, Inc, 206
- Robert. E Slavin. 2010. *Cooperative Learning Teori, Riset dan Praktik*. Bandung:Nusa Media
- Slamet 2010.*Belajar Dan Factor-Faktor Yang Mempengaruhinya*.Jakarta:Rineka Cipta.