

Penerapan Machine Learning dengan Algoritma C4.5 Untuk Memprediksi Kesesuaian Gaya Belajar Siswa Madrasah Ibtidaiyah

Wiwik Widianti¹, Fatimatus syahro², Dinda Lailatul Qomariyah³, Novi Fitriani⁴

Universitas Nurul Jadid, Probolinggo, Indonesia^{1,2,3,4}

{wiwikwidianti1302@gmail.com¹, zahro3311@gmail.com², dindailatulkomariyah@gmail.com³, fitriani29nop@gmail.com⁴}

Abstrak. Data Mining adalah proses mengekstraksi dan menemukan pola dalam kumpulan data besar yang melibatkan metode dalam proses pembelajaran mesin, statistik, dan sistem basis data. MI Raudlatul Mutaallimin merupakan sekolah dasar yang terletak di desa Pandean kecamatan paiton. Pada proses pembelajaran di MI Raudlatul Mutaallimin, masih banyak murid yang tidak fokus dan tidak memiliki kemauan untuk belajar, kemungkinan terbesar adalah ketidakcocokan gaya belajar dan metode belajar guru tersebut. Oleh karenanya, penulis melakukan analisis data mining terhadap data murid dan kepribadiannya agar data tersebut dapat berubah menjadi informasi yang bermanfaat bagi lembaga. Informasi ini bisa membantu untuk meningkatkan proses belajar mengajar di MI Raudlatul Mutaallimin. Penulis menggunakan hasil data quisioner terhadap siswa kelas 4, 5 dan 6 sebanyak 165 data siswa-siswi yang telah penulis sajikan. Dalam proses analisis ini, penulis menggunakan tools Rapiminer Studio. Metode yang digunakan decision tree dengan algoritma C4.5. Hasil prediksi menggunakan algoritma decision tree C4.5 dengan hasil 8 rules. Persentasi hasil akurasi decision tree dengan menggunakan 10 Fold Cross Validation membuktikan bahwa tingkat akurasinya sebesar 81.18%. dengan nilai class precission untuk prediksi kinestetik sebesar 85.43%, prediksi visual sebesar 33.33%, dan prediksi auditorial sebesar 0.00%, hasil 5 Fold Cross Validation akurasinya sebesar 82.27%. dengan nilai class precission untuk prediksi kinestetik sebesar 86.09%, prediksi visual sebesar 41.67%, dan prediksi auditorial sebesar 0.00%. Dari hasil perbandingan membuktikan implementasi dengan model 5 Fold Cross Validation lebih baik.

Katakunci: *Data Mining; Algoritma C 4.5; Gaya Belajar Siswa*

Pendahuluan

Proses belajar mengajar yang efektif tentunya sangat menentukan mutu serta kualitas dalam bidang pendidikan. Beberapa cara yang dapat digunakan untuk meningkatkan proses belajar mengajar dalam suatu institusi pendidikan, seperti menambah atau meningkatkan fasilitas belajar, peningkatan kualitas tenaga pengajar (guru), dan pembagian kelas yang terjadwal dengan baik. Permasalahan yang sering terjadi dalam pembagian kelas adalah perbedaan gaya belajar pada masing-masing siswa sehingga tenaga pengajar (guru) kesulitan dalam memilih metode yang tepat untuk digunakan (Gani, 2018).

Secara umum gaya belajar manusia dibedakan ke dalam tiga kelompok besar, yaitu gaya belajar visual, gaya belajar auditorial dan gaya belajar kinestetik. Dimana gaya belajar visual adalah gaya belajar dengan cara melihat, mengamati, dan memandang. Gaya belajar auditorial adalah gaya belajar dengan mendengar. Dan gaya belajar kinestetik adalah gaya belajar dengan cara bergerak, berkerja, dan menyentuh. Peningkatan kualitas belajar tentunya harus dimulai di saat peserta didik berada pada jenjang SD (Sekolah Dasar). Pada sistem pembagian kelas di SD Negeri Sidodadi Paiton masih diacak tanpa berdasarkan kesesuaian gaya belajar, sehingga hal tersebut membuat proses belajar mengajar menjadi tidak efektif (Wahyuni, 2017).

Dari hasil pengamatan dan observasi peneliti di SD Negeri Sidodadi Paiton, masih banyak siswa-siswi yang bermain-main atau tidak tertarik saat mengikuti proses belajar mengajar. Hal ini membuktikan bahwa siswa-siswi belum termotivasi dalam kegiatan belajar. Gaya belajar pun juga bisa mempengaruhi pribadi peserta didik dalam belajarnya, dapat dikatakan peserta didik termotivasi secara ilmiah membuat proses belajar peserta didik lebih efektif. Dengan munculnya motivasi, peserta didik akan memiliki keinginan untuk belajar (Purwaningsih, 2021).

Data Mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar. Data mining bisa disebut knowledge discovery adalah proses pengambilan pola pada data yang akan di proses lalu output tersebut berupa informasi yang sangat penting (Romero & Ventura, 2013).

Decision tree adalah alat pendukung keputusan yang menggunakan model keputusan seperti pohon dan kemungkinan konsekuensinya, termasuk hasil peristiwa kebetulan, biaya sumber daya, dan utilitas. Algoritma decision tree adalah salah satu algoritma pembelajaran mesin dengan tipe diawasi dan populer yang digunakan untuk klasifikasi. Algoritma ini menghasilkan hasil sebagai hasil yang dioptimalkan berdasarkan struktur pohon dengan kondisi atau aturan (Safavian & Landgrebe, 1991).

Algoritma C4.5 digunakan dalam Data Mining sebagai Decision Tree Classifier yang dapat digunakan untuk menghasilkan keputusan, berdasarkan sampel data tertentu (prediktor univariat atau multivariat). Algoritma C4.5 lebih efektif, hasil ketepatan prediksi dan nilai kesalahan (error rate) lebih baik dari ID3. Pengukuran kinerja yang dilakukan

menggunakan sekelompok data uji untuk mengetahui prosentase precision, recall dan accuracy, menunjukkan bahwa algoritma C4.5 memiliki tingkat akurasi yang lebih tinggi dari pada algoritma ID3. Berdasarkan kasus diatas pada penelitian ini menggunakan algoritma C4.5 dalam melakukan klasifikasi Kesesuaian Gaya Belajar Siswa SD dengan tujuan membantu pihak sekolah dalam memotifasi siswa agar peserta didik akan memiliki keinginan untuk belajar dengan baik(Mardi, 2017).

Metode

Kerangka kerja penelitian dibuat agar mempermudah pencapaian hasil penelitian, dapat berjalan sesuai dengan yang diharapkan. Adapun kerangka kerja penelitian yang digunakan dapat dilihat pada gambar 1:



Gambar 1. Kerangka Penelitian

[66]

1. Perumusan

Masalah Pada tahap ini dilakukan peninjauan pada objek penelitian untuk melakukan observasi langsung sehingga menemukan permasalahan-permasalahan yang ada pada MI Raudlatul Mutaallimin Pandean.

2. Studi Literatur

Pada tahap ini penulis melakukan pencarian terhadap landasan-landasan teori yang diperoleh dari berbagai sumber seperti buku, jurnal ilmiah (jurnal nasional dan jurnal internasional) dan juga referensi lainnya guna untuk melengkapi penelitian baik mengenai konsep dan teori sehingga memiliki acuan yang baik julas dan relevan dengan topik yang di angkat.

3. Pengumpulan Data

Pada tahap ini penulis melakukan pengumpulan data dan informasi yang dibutuhkan dalam penelitian. Pengumpulan data ini dilakukan dengan beberapa metode yaitu:

a. Pengamatan (Observasi)

Metode ini dilakukan dengan pengamatan secara langsung pada objek yang akan diteliti. Dalam hal ini penulis mengamati proses belajar mengajar pada pada lembaga MI Raudlatul Mutaallimin.

- b. Wawancara

Pada metode ini penulis melakukan wawancara singkat secara langsung dengan kepala MI Raudlatul Mutaallimin Pandean untuk memperoleh data dan informasi mengenai data siswaswi yang penulis butuhkan.
4. Perhitungan dan Analisis Decision Tree dengan Algoritma C4.5

Pada tahapan ini penulis melakukan perhitungan decision tree dengan menggunakan algoritma C4.5 dengan menghitung nilai gain dan entropy dari data-data yang telah dinormalisasi dan digunakan untuk membuat hasil decision tree.
5. Hasil Analisis

Pada tahapan ini penulis menganalisis hasil dari perhitungan algoritma C4.5 dan pembuatan decision tree. Hasil yang didapat bisa digunakan untuk memprediksi gaya belajar yang cocok untuk masing-masing siswa-siswi MI Raudlatul Mutaallimin Pandean dengan melihat dari hasil decision tree.
6. Penulisan Laporan dan Penarikan Kesimpulan

Pada tahapan ini dilakukan pembuatan laporan yang disusun berdasarkan hasil penelitian dengan menggunakan teknik pengumpulan data primer dan data sekunder, sehingga menjadi laporan penelitian yang sesuai dengan analisis yang sedang diteliti.

Hasil dan Pembahasan

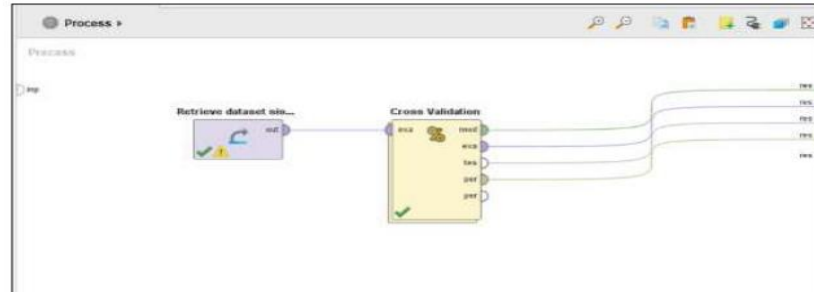
Berdasarkan hasil dari observasi dan wawancara yang dilakukan, penulis mendapatkan data siswa-siswi di MI Raudlatul Mutaallimin. Data yang diperoleh sebanyak 165 data quisioner siswa kelas 4, 5 dan 6. Berikut ini rekap data hasil quisioner yang disajikan dalam bentuk tabel. Atribut yang digunakan pada penelitian ini mencakup Jenis Kelamin, Hoby, Kelas Favorit, Kemampuan, Metode Belajar dan Gaya. Berikut tabel dataset yang akan digunakan dalam implementasi Algoritma C4.5.

COREAI Jurnal Kecerdasan Buatan, Komputasi dan Teknologi Informasi
 Vol.3 No.2 Tahun 2022 : 65-72

Tabel 1. Dataset hasil *Quisioner* Siswa

No	Jenis Kelamin	Hoby	Kelas Favorit	Kemampuan	Metode Belajar	Gaya Belajar
1	LAKI - LAKI	OLAHRAGA	MUSIK	BERNYAYI	MEMBACA CATATAN	KINESTETIK
2	LAKI - LAKI	KESENIAN	OLAHRAGA	OLAHRAGA	MENULIS ULANG	VISUAL
3	PEREMPUAN	OLAHRAGA	OLAHRAGA	OLAHRAGA	TANYA JAWAB	VISUAL
4	LAKI - LAKI	MUSIK	KESENIAN	MENGGAMBAR	MEMBACA CATATAN	AUDITORIAL
5	LAKI - LAKI	MUSIK	OLAHRAGA	BERNYAYI	MEMBACA CATATAN	AUDITORIAL
6	PEREMPUAN	OLAHRAGA	OLAHRAGA	OLAHRAGA	MENULIS ULANG	VISUAL
7	PEREMPUAN	KESENIAN	OLAHRAGA	BERNYAYI	TANYA JAWAB	KINESTETIK
8	PEREMPUAN	KESENIAN	MUSIK	BERNYAYI	MEMBACA CATATAN	VISUAL
...
162	LAKI - LAKI	OLAHRAGA	OLAHRAGA	MENGGAMBAR	MENULIS ULANG	VISUAL
163	LAKI - LAKI	OLAHRAGA	MUSIK	BERNYAYI	MENULIS ULANG	AUDITORIAL
164	PEREMPUAN	MUSIK	KESENIAN	BERNYAYI	MEMBACA CATATAN	KINESTETIK
165	LAKI - LAKI	OLAHRAGA	KESENIAN	MENGGAMBAR	MENULIS ULANG	VISUAL

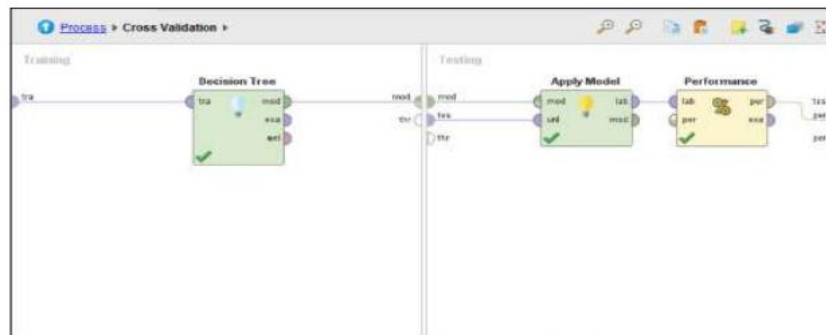
Pada proses penerapannya pada gambar 2 di bawah, aplikasi yang digunakan adalah aplikasi *Rapidminer studio* dengan pengujian menggunakan *Cross Validation* dengan proses *10-fold cross validation* dan *5-fold cross validation* untuk algoritma yang digunakan adalah *decision tree C4.5*, untuk data *training* dan data *testing default 75%* untuk data *training* dan 25% untuk data *testing*.



Gambar 2. Desain Implementasi proses algoritma *decision tree C4.5*

Proses selanjutnya menguji model Operator ini menerapkan model pada *ExampleSet*. Sedangkan untuk operator performance digunakan untuk evaluasi kinerja dan memberikan daftar nilai kriteria kinerja. Kriteria kinerja ini secara otomatis ditentukan agar sesuai dengan jenis tugas pembelajaran berikut implementasinya pada gambar 3.

COREAI Jurnal Kecerdasan Buatan, Komputasi dan Teknologi Informasi (2022)



Gambar 3. Desain Implementasi yang ada di dalam *Cross Validation*

Hasil pengujian algoritma *decision tree C4.5* dengan data *training* sebanyak 75% dan data *testing* sebanyak 25% *modeling* yang digunakan *10 Fold Cross Validation* membuktikan bahwa tingkat akurasi sebesar 81.18%. dengan nilai *class precision* untuk prediksi kinestetik sebesar 85.43%, prediksi visual sebesar 33.33%, dan prediksi auditorial sebesar 0.00%, hasil *5 Fold Cross Validation* akurasi sebesar 82.27%. dengan nilai *class precision* untuk prediksi kinestetik sebesar 86.09%, prediksi visual sebesar 41.67%, dan prediksi auditorial sebesar 0.00%, detainya dapat dilihat pada table 2.

Tabel 2. Perbandingan Evaluasi Akurasi

Model Evaluasi	Accuracy	Class Precision		
		pred. KINESTETIK	pred. VISUAL	pred. AUDITORIAL
5 Fold Cross Validation	82.27%	86.09%	41.67%	0.00%
10 Fold Cross Validation	81.18%	85.43%	33.33%	0.00%

Selain dari tabel *Confusion Matrix* di atas, terdapat juga hasil dari *performance vector* untuk metode *decision tree* C4.5. *Performance vector* sendiri merupakan bentuk dari deskripsi tabel hasil dari analisis. Hanya saja terdapat tambahan kappa pada *performance vector* metode *decision tree*, detailnya bisa dilihat pada gambar 4 di bawah ini.

COREAI Jurnal Kecerdasan Buatan, Komputasi dan Teknologi Informasi
 Vol.3 No.2 Tahun 2022 : 65-72

```

PerformanceVector

PerformanceVector:
accuracy: 81.18% +/- 7.08% (micro average: 81.10%)
ConfusionMatrix:
True:  KINESTETIK      VISUAL  AUDITORIAL
KINESTETIK:      129      13      9
VISUAL: 7         4        1
AUDITORIAL:      1         0        0
kappa: 0.145 +/- 0.283 (micro average: 0.152)
ConfusionMatrix:
True:  KINESTETIK      VISUAL  AUDITORIAL
KINESTETIK:      129      13      9
VISUAL: 7         4        1
AUDITORIAL:      1         0        0
    
```

Gambar 4. Berikan hasil *performance vector*

Hasil analisis dengan *Tools Rapidminer studio* dengan dataset sebanyak 165 kemudian dibagi kedalam data training dan data testing yang mana untuk data *training* 75% sedangkan untuk data *testing* 25% diambil secara random, berikut hasil dalam bentuk *decision tree* bisa dilihat pada gmabra 5.



Gambar 5. Hasil Decision Tree C4.5 menggunakan tools Rapidminer studio

Decision Tree pada gambar 5, diperoleh dengan menggunakan tools Rapidminer Studio dengan algoritma C4.5 menggunakan model 5 Fold Cross Validation dan 10 Fold Cross Validation. Rules seperti yang telah di jelaskan sebelumnya. Rules tersebut bisa dilihat pada gambar 6.

COREAI Jurnal Kecerdasan Buatan, Komputasi dan Teknologi Informasi (2022)

```

Tree
Jenis Kelamin = LAKI - LAKI
| Hoby = KESENIAN
| | Metode Belajar = MENULIS ULANG: VISUAL (KINESTETIK=0, VISUAL=3, AUDITORIAL=1)
| | Metode Belajar = TANYA JAWAB: AUDITORIAL (KINESTETIK=1, VISUAL=0, AUDITORIAL=2)
| Hoby = MUSIK: KINESTETIK (KINESTETIK=49, VISUAL=0, AUDITORIAL=6)
| Hoby = OLAHRAGA: KINESTETIK (KINESTETIK=30, VISUAL=3, AUDITORIAL=1)
Jenis Kelamin = PEREMPUAN
| Hoby = KESENIAN: KINESTETIK (KINESTETIK=23, VISUAL=4, AUDITORIAL=0)
| Hoby = MUSIK: KINESTETIK (KINESTETIK=28, VISUAL=1, AUDITORIAL=0)
| Hoby = OLAHRAGA
| | Kelas Favorit = KESENIAN: KINESTETIK (KINESTETIK=2, VISUAL=0, AUDITORIAL=0)
| | Kelas Favorit = OLAHRAGA: VISUAL (KINESTETIK=4, VISUAL=6, AUDITORIAL=0)
    
```

Gambar 6. Rules Decision Tree Menggunakan Algoritma C4.5

Kesimpulan dan Saran

Dari pembahasan di atas dapat diambil kesimpulan bahwa hasil penelitian implementasi Algoritma C4.5 untuk memprediksi Kesesuaian Gaya Belajar Siswa Sekolah Dasar ini menggunakan data siswa SD Negeri Sidodadi Paiton tahun ajaran 2020/2021 dengan jumlah data siswa 165 dan menggunakan 6 atribut untuk proses prediksi yaitu, Jenis Kelamin, Hoby, Kelas Favorit, Kemampuan, Metode Belajar dan Gaya Belajar. Hasil

prediksi Gaya Belajar terdapat 3 jenis Gaya Belajar yaitu, Visual, Auditorial dan Kinestetik. Hasil prediksi menggunakan algoritma decision tree C4.5 dengan hasil 8 rules prediksi gaya belajar siswa. Dimana hasil dari rules decision tree tersebut dapat digunakan untuk memprediksi masing-masing gaya belajar siswa. Persentasi hasil akurasi decision tree dengan menggunakan 10 Fold Cross Validation membuktikan bahwa tingkat akurasinya sebesar 81.18%. dengan nilai class precision untuk prediksi kinestetik sebesar 85.43%, prediksi visual sebesar 33.33%, dan prediksi auditorial sebesar 0.00%, hasil 5 Fold Cross Validation akurasinya sebesar 82.27%. dengan nilai class precision untuk prediksi kinestetik sebesar 86.09%, prediksi visual sebesar 41.67%, dan prediksi auditorial sebesar 0.00%. Dari hasil perbandingan membuktikan implementasi dengan model 5 Fold Cross Validation lebih baik.

Daftar Pustaka

- Gani, A. A. (2018). Interaksi antara pemanfaatan media pembelajaran dan gaya belajar terhadap hasil belajar IPS terpadu. *CIVICUS: Pendidikan-Penelitian-Pengabdian Pendidikan Pancasila Dan Kewarganegaraan*, 6(2), 83–87.
- Mardi, Y. (2017). Data Mining: Klasifikasi Menggunakan Algoritma C4. 5. *Jurnal Edik Informatika Penelitian Bidang Komputer Sains Dan Pendidikan Informatika*, 2(2), 213–219.
- Purwaningsih, H. (2021). Peran Guru Bimbingan dan Konseling dalam melayani peserta didik di masa pandemi covid-19. *Educational: Jurnal Inovasi Pendidikan & Pengajaran*, 1(1), 36–44.
- Romero, C., & Ventura, S. (2013). Data mining in education. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 3(1), 12–27.
- Safavian, S. R., & Landgrebe, D. (1991). A survey of decision tree classifier methodology. *IEEE Transactions on Systems, Man, and Cybernetics*, 21(3), 660–674.
- Wahyuni, Y. (2017). Identifikasi gaya belajar (visual, auditorial, kinestetik) mahasiswa pendidikan matematika universitas bung hatta. *JPPM (Jurnal Penelitian Dan Pembelajaran Matematika)*, 10(2).