

Kajian Klasifikasi Data Mining IQ Siswa SMA Berdasarkan Hasil Intelligence Structure Test Dengan Menggunakan Metode Naive Bayes

Des Suryani¹, Ause Labellapansa², M Luthfan Shiddiqie³, Ahmad Hidayat⁴

Jurusan Teknik Informatika Universitas Islam Riau

Jl. Kaharuddin Nasution No. 113 Pekanbaru 28284

Telp (0761) 674674, fax (0761) 674834

des.suryani@eng.uir.ac.id¹, ause.labella@eng.uir.ac.id²,

muhammadluthfan28@student.uir.ac.id³, ahmadhidayat@psy.uir.ac.id⁴

Abstract. SMA Negeri 1 Padang menyelenggarakan tes psikologi dengan menggunakan Intelligence Structure Test (IST) sebagai alat tes psikologi untuk menentukan klasifikasi IQ siswa. Klasifikasi IQ siswa ini dilayani oleh seorang psikolog. Hasil klasifikasi IQ siswa ini belum dilakukan evaluasi untuk mendapatkan pengetahuan dari kumpulan data tersebut. Untuk itu perlu dilakukan penelitian dalam melihat peluang di masa yang akan datang untuk membantu psikolog melakukan klasifikasi IQ siswa. Dalam hal ini menerapkan konsep data mining dengan metode Naive Bayes dan menggunakan skala David Wechsler sebagai instrumen penelitian. Berdasarkan pengujian 100 data uji pada 232 dan 332 data uji, penelitian ini mampu mengklasifikasikan IQ peserta dengan tingkat akurasi yang lebih dari 90%. Dengan adanya sistem ini maka klasifikasi IQ dapat dilakukan lebih cepat dan akurat.

Keywords: IQ, IST, Klasifikasi, Datamining

1 Pendahuluan

Intelligence Structure Test (IST) adalah salah satu alat tes inteligensi yang dikembangkan pada tahun 1953 oleh Rudolf Amthauer di Frankfurt Main Jerman dan telah diadaptasi di Indonesia. *Alat tes ini* merupakan suatu bentuk yang terdiri dari bagian-bagian yang saling berhubungan dan mempunyai makna.

Subtes yang dimuat dalam IST dan sering digunakan ada 9 (Sembilan) yaitu melengkapi kalimat (*Satzerganzung/SE*), melengkapi kata-kata (*Wortauswahl/WA*), persamaan kata (*Analogien/AN*), sifat yang dimiliki bersama (*Gemeinsamkeiten/GE*), kemampuan berhitung (*Rechenaufgaben/RA*), deret angka (*Zahlenreihen/ZR*), memilih bentuk (*Figurenauswahl/FA*), latihan balok (*Wurfelaufgaben/WU*), dan latihan simbol (*Merkaufgaben/ME*).

SMA Negeri 1 Padang menyelenggarakan tes psikologi dengan menggunakan tes IST sebagai alat tes psikologi untuk menentukan klasifikasi IQ siswa. Tes IST yang digunakan memiliki 10 kategori penilaian antara lain umur, SE, WA, AN, GE, RA, SR, FA, WU dan ME dengan menentukan skala klasifikasi IQ siswa menggunakan instrumen David Wechsler yang terdiri dari 7 (tujuh) kategori antara lain sangat superior, superior, diatas rata-rata, rata-rata, di bawah rata-rata, *borderline*, dan *intellectual deficient*. Klasifikasi IQ siswa ditangani oleh psikolog dan membutuhkan waktu sekitar 72 menit per siswa.

Penelitian untuk data yang sama telah dilakukan dengan menggunakan algoritma C4.5 dan hasil akurasi klasifikasi penggolongan IQ siswa diperoleh 84%¹. Sedangkan penelitian ini mencoba membangun sistem yang dapat membantu psikolog dalam melakukan klasifikasi IQ siswa menggunakan metode klasifikasi dengan algoritma *Naive Bayes*. Penggunaan metode Naive Bayes dapat membantu dalam penelitian ini karena data yang diolah berupa data kuantitatif dan tidak memerlukan data latih dalam jumlah yang besar.

2 Dasar Teori

2.1 Intelligence Quotient (IQ)

Inteligensi merupakan kemampuan mental yang menggabungkan proses berpikir secara rasional. Intelegensia tidak dapat diamati secara langsung, melainkan kesimpulan dari berbagai tindakan nyata yang merupakan manifestasi dari proses berpikir rasional. Sementara itu IQ adalah skor yang diperoleh dari sebuah alat tes kecerdasan. Tes tersebut hanya memberikan sedikit indikasi mengenai taraf kecerdasan seseorang, namun tidak menggambarkan kecerdasan seseorang secara keseluruhan².

Tingkat inteligensi merupakan keseluruhan kecakapan yang dimiliki seseorang sehingga dapat bertindak dan berpikir secara terarah dan baik. Tingkat inteligensi seseorang khususnya siswa dapat diukur melalui tes IQ³.

2.2 Intelligence Structure Test (IST)

Intelligence Structure Test (IST) merupakan alat tes inteligensi yang dikembangkan oleh Rudolf Amthaeur di Frankfurt Main Jerman pada tahun 1953 dan telah diadaptasi di Indonesia. IST berdasarkan pada teori inteligensi yang menyatakan bahwa inteligensi merupakan suatu *gestalt* yang terdiri dari bagian-bagian yang saling berhubungan secara bermakna⁴.

Norma IST tersedia untuk umur 12 tahun sampai umur 60 tahun dengan pembagian kelompok umur sebagai berikut 12, 13, 14, 15, 16, 17, 18, 19-20, 21-25, 26-30, 31-35, 36-40, 41-45, 46-50 dan 51-60 tahun. Norma IST per kelompok umur tersedia untuk masing-masing subtes serta norma untuk mendapatkan IQ. Norma IST disusun dengan mengubah skor mentah menjadi *Standardized Score* (SS) atau skor baku dengan rata-rata (*mean*, disingkat M) 100 dan simpang baku (*Standard Deviation*, disingkat S) 10. Penggunaan tabel norma per kelompok umur adalah sebagai berikut :

1. Cocokkan umur peserta tes dengan tabel norma kelompok umur.
2. Dapatkan skor baku atau norma SS dengan mencocokkan skor mentah (RS) peserta tes pada tabel norma subtes yang sesuai.
3. Pada tabel dibagian atas lembar jawaban halaman pertama, tuliskan skor mentah (RS) dan skor baku (SS) untuk subtes yang sesuai.
4. Menggolongkan IQ per subtes berdasarkan penggolongan Wechsler, gunakan tabel konversi SS IST ke IQ Wechsler dan persentil (semua umur) untuk mengkonversi SS ke dalam IQ Wechsler. Lalu gunakan tabel 1 untuk penggolongan IQ berdasarkan Skala David Wechsler untuk menentukan kategori IQ.

Tabel 1. Penggolongan IQ berdasarkan skala David Wechsler

| <i>SS IST</i> | <i>IQ</i> | <i>Kategori</i> |
|---------------|-----------|------------------------|
| ≥ 120 | ≥ 130 | Sangat Superior |
| 113 – 119 | 120 – 129 | Superior |
| 107 – 112 | 110 – 119 | Di atas rata-rata |
| 93 – 106 | 90 – 109 | Rata-rata |
| 87 – 92 | 80 – 89 | Di bawah rata-rata |
| 80 – 86 | 70 – 79 | Lambat Belajar |
| ≤ 69 | ≤ 69 | Keterbelakangan Mental |

2.3 Data Mining

Data mining saat ini banyak diaplikasikan dalam berbagai bidang, salah satunya di bidang psikologi. Jumlah data yang besar pada bidang psikologi seperti data hasil test IQ siswa SMA dapat dianalisis untuk membentuk model pengetahuan yang akan digunakan psikolog. Dengan meningkatnya jumlah data pada beberapa tahun terakhir baik dalam bentuk data terstruktur maupun tidak terstruktur perlu dilakukan analisis untuk mendapatkan keputusan-keputusan strategis⁴.

2.4 Algoritma Naive Bayes

Dalam penelitian ini menggunakan Algoritma Naïve Bayes yang merupakan salah satu algoritma yang digunakan dalam teknik klasifikasi⁵. Adapun alur dari metode Naive Bayes adalah sebagai berikut :

- a. Baca data training
- b. Hitung jumlah dan probabilitas, namun apabila data numerik maka gunakan persamaan berikut:

$$P(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (1)$$

Keterangan :
 σ = Standar deviasi
 μ = mean

c. Hitung nilai likelihood dengan persamaan berikut :

$$Likelihood (X)_i = \sum P(x) \quad (2)$$

d. Hitung nilai probabilitas prior dengan persamaan berikut :

$$Probabilitas Prior (X)_i = \frac{Likelihood(x)_i}{\sum Likelihood(x)} \quad (3)$$

2.5 Confusion Matrix

Pada konsep data mining, untuk menghitung akurasi pada teknik klasifikasi dapat menggunakan metode *Confusion matrix* yang merupakan tabel yang sering digunakan untuk menggambarkan kinerja model klasifikasi (*classifier*) pada satu set data uji yang nilai sebenarnya diketahui. Ini merupakan kinerja dari suatu algoritma. Presisi atau *confidence* adalah proporsi kasus yang diprediksi positif yang juga positif benar pada data yang sebenarnya. Recall atau *sensitivity* adalah proporsi kasus positif yang sebenarnya yang diprediksi positif secara benar⁶.

Tabel 2. Model *confusion matrix*

| Aktual | Hasil Klasifikasi | |
|--------|--------------------|--------------------|
| | + | - |
| + | True Positive (A) | True Negative (B) |
| - | False Positive (C) | False Negative (D) |

Perhitungan akurasi dengan tabel *confusion matrix* adalah sebagai berikut:

$$\text{Akurasi} = \frac{(A+D)}{(A+B+C+D)} * 100\% \quad (4)$$

3 Pembahasan

3.1 Perhitungan Metode Naive Bayes

Penelitian ini bertempat di Sumatera Barat yaitu di kota Padang tepatnya di SMA Negeri 1 Padang. Data training yang digunakan dalam penelitian ini adalah data hasil tes IQ siswa SMA Negeri 1 Padang tahun 2017 sebanyak 332 data sedangkan untuk pengujian data, data yang akan diuji terdiri dari 100 data uji yang dilakukan terhadap 232 data training dan 332 data training. Diketahui hasil tes IQ seorang siswa yang dapat dilihat pada tabel 3.

Tabel 3. Contoh data uji

| Umur | WA | AN | ME | GE | RA | WU | ZR | FA | SE | IQ |
|------|-----|-----|-----|-----|-----|-----|-----|-----|----|-----|
| 13 | 105 | 117 | 110 | 104 | 134 | 106 | 101 | 110 | 96 | 125 |

Langkah pertama adalah menghitung nilai probabilitas setiap kelas yang dapat dilihat pada tabel 4.

Tabel 4. Nilai probabilitas semua kelas

| Kelas | Jumlah | Probabilitas |
|------------------|--------|--------------|
| Sangat Superior | 13 | 0,043046358 |
| Superior | 50 | 0,165562914 |
| Diatas Rata-Rata | 100 | 0,331125828 |
| Rata-Rata | 139 | 0,460264901 |

Langkah selanjutnya adalah menghitung nilai probabilitas pada setiap kriteria pada masing-masing kelas. Untuk menghitung nilai probabilitas pada setiap kriteria pada masing-masing kelas diperlukan menghitung *mean* dan standar deviasi dari setiap kriteria terhadap masing-masing kelas terlebih dahulu seperti pada tabel 5 dan 6.

Tabel 5. Standar deviasi setiap kriteria pada masing-masing kelas

| Klasifikasi | Sangat Superior | Superior | Diatas Rata-Rata | Rata-Rata |
|-------------|-----------------|----------|------------------|-----------|
| Umur | 0,421 | 0,480 | 0,518 | 0,479 |
| WA | 7,202 | 6,961 | 7,076 | 8,538 |
| AN | 6,766 | 7,275 | 6,650 | 7,025 |
| ME | 3,988 | 6,934 | 9,683 | 10,129 |
| GE | 6,191 | 5,211 | 5,975 | 6,582 |
| RA | 10,453 | 10,078 | 8,803 | 7,235 |
| WU | 8,837 | 9,958 | 10,243 | 12,840 |
| ZR | 3,013 | 9,999 | 8,728 | 8,721 |
| FA | 7,323 | 6,758 | 7,367 | 7,255 |
| SE | 3,585 | 16,345 | 12,414 | 7,369 |
| IQ | 3,231 | 3,169 | 2,919 | 5,755 |

Tabel 6. Mean setiap kriteria pada masing-masing kelas

| Klasifikasi | Sangat Superior | Superior | Diatas Rata-Rata | Rata-Rata |
|-------------|-----------------|----------|------------------|-----------|
| Umur | 14,23 | 14,64 | 14,46 | 14,85 |
| WA | 123,23 | 118,78 | 115,24 | 109,12 |
| AN | 116,62 | 111,8 | 109 | 101,52 |
| ME | 126,31 | 123,2 | 118,13 | 112,25 |
| GE | 113,77 | 107,74 | 107,62 | 103,22 |
| RA | 102,23 | 100,28 | 93,96 | 86,78 |
| WU | 120,46 | 110,14 | 105,8 | 97,62 |
| ZR | 127,00 | 117,02 | 107,49 | 99,43 |
| FA | 108,38 | 103,92 | 101,47 | 97,79 |
| SE | 111,38 | 104,18 | 102,11 | 97,35 |
| IQ | 138,15 | 124,28 | 114,61 | 100,61 |

Setelah *mean* dan standar deviasi dari setiap kriteria pada masing-masing kelas diperoleh maka langkah selanjutnya adalah menghitung nilai probabilitas kriteria pada masing-masing kelas dengan menggunakan persamaan (1) seperti pada contoh berikut:

$$\begin{aligned}
 P(x) &= \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \\
 &= \frac{1}{\sqrt{2\pi} * 0,421} e^{-\frac{(13-14,23)^2}{2*0,421^2}} \\
 &= 0,0133
 \end{aligned}$$

Hasil seluruh perhitungan nilai probabilitas kriteria pada masing-masing kelas dapat dilihat pada tabel 7.

Tabel 7. Nilai probabilitas kriteria pada masing-masing kelas

| Klasifikasi | Sangat Superior | Superior | Diatas Rata-Rata | Rata-Rata |
|-------------|-----------------|----------|------------------|-----------|
| Umur=13 | 0,0133 | 0,0024 | 0,0145 | 0,0005 |
| WA=105 | 0,0022 | 0,0081 | 0,0198 | 0,0416 |
| AN=117 | 0,0589 | 0,0425 | 0,0291 | 0,0050 |
| ME=110 | 0,0000 | 0,0094 | 0,0290 | 0,0384 |
| GE=104 | 0,0186 | 0,0592 | 0,0556 | 0,0602 |
| RA=134 | 0,0004 | 0,0001 | 0,0000 | 0,0000 |
| WU=106 | 0,0118 | 0,0368 | 0,0390 | 0,0251 |
| ZR=101 | 0,0000 | 0,0111 | 0,0347 | 0,0450 |
| FA=110 | 0,0532 | 0,0394 | 0,0277 | 0,0134 |
| SE=96 | 0,0000 | 0,0215 | 0,0285 | 0,0533 |
| IQ=125 | 0,0000 | 0,1227 | 0,0002 | 0,0000 |

Langkah selanjutnya adalah menentukan nilai likelihood pada setiap kelas dengan menggunakan persamaan (2). Likelihood adalah peluang munculnya karakteristik sampel pada kelas yang ada. Berikut adalah contoh perhitungan likelihood :

$$\begin{aligned}
 \text{Likelihood } (X)_i &= \sum P(x) \\
 &= 0,0133*0,0022*0,0589*....*0,0430 \\
 &= 2,41036E-47
 \end{aligned}$$

Hasil perhitungan likelihood dapat dilihat pada tabel 8.

Tabel 8. Nilai likelihood pada setiap kelas

| Klasifikasi | Likelihood |
|------------------|-------------|
| Sangat Superior | 2,41036E-47 |
| Superior | 4,76402E-22 |
| Diatas Rata-Rata | 1,67954E-24 |
| Rata-Rata | 2,31181E-32 |

Langkah terakhir adalah menghitung nilai probabilitas prior untuk menentukan kelas data yang baru dengan persamaan (3). Probabilitas prior adalah peluang munculnya masing-masing kelas yang ada. Berikut adalah contoh perhitungan probabilitas prior:

$$\begin{aligned}
 \text{Probabilitas Prior } (X)_i &= \frac{\text{Likelihood}(x)_i}{\sum \text{Likelihood}(x)} \\
 &= \frac{2,41036E - 47}{2,41036E - 47 + \dots + 2,31181E - 32} \\
 &= 0,000000000000000000000000050417
 \end{aligned}$$

- (2018).
4. Abbamonte, G. Foreword. *New Horizons a Data-Driven Econ. A Roadmap Usage Exploit. Big Data Eur.* v–vi (2016). doi:10.1007/978-3-319-21569-3
 5. Qurnia, I. L. *et al.* CLASSIFICATION OF DIABETES DISEASE USING Case Study : **1**, 147–151 (2016).
 6. Novakovic, J., Veljovi, A., Iiic, S., Papic, Z. & Tomovic, M. Evaluation of Classification Models in Machine Learning. *Theory Appl. Math. Comput. Sci.* **7**, 39–46 (2017).