

KLASIFIKASI TINGKATAN LEVEL KUMON DENGAN MENGGUNAKAN METODE *NAÏVE BAYES CLASSIFIER*

Siti Aisyah^{1*}, Nina Valentika², Aden³
Program Studi Matematika, FMIPA Universitas Pamulang^{1,2,3}
nauraisyah67@gmail.com¹, dosen02339@unpam.ac.id², dosen00527@unpam.ac.id³

Abstrak-Kumon merupakan bimbingan belajar yang menerapkan sistem pembelajaran berdasarkan level yang disesuaikan dengan kemampuan individu, bertujuan untuk mengoptimalkan kemampuan serta kecerdasan setiap individu. Dengan metode pembelajaran kumon banyak siswa yang dapat belajar dengan level di atas tingkatan kelas, di bawah tingkatan kelas, dan setara tingkatan kelas. Dalam menentukan tingkatan level kumon biasanya baru terlihat pada siswa saat mengikuti kelas kumon selama 6 bulan sampai dengan 1 tahun. Namun dengan menggunakan metode *Naïve Bayes Classifier* akan mudah dan cepat dalam memprediksi tingkatan level kumon pada siswa dengan melihat faktor-faktor yang mempengaruhi tingkatan level siswa seperti motivasi belajar, minat belajar, kebiasaan belajar, dan jenis kelamin siswa.

Dari penelitian yang telah dilakukan berdasarkan hasil pengujian *confusion matrix*, diperoleh nilai *accuracy* sebesar 70%, *class precision* sebesar 75% kategori di atas tingkatan kelas, *class precision* sebesar 100% kategori di bawah tingkatan kelas, *class precision* sebesar 60% kategori setara tingkatan kelas, *class recall* sebesar 100% kategori di atas tingkatan kelas, *class recall* sebesar 25% kategori di bawah tingkatan kelas, dan *class recall* sebesar 100% kategori setara tingkatan kelas. Sehingga dapat dinyatakan *Algoritma Naïve Bayes Classifier* dapat digunakan untuk memprediksi klasifikasi tingkatan level kumon. Namun diharapkan penelitian berikutnya menggunakan atribut lain agar mendapatkan algoritma yang lebih akurat.

Kata Kunci: *Naïve Bayes Classifier, Tingkatan Level Kumon.*

I. PENDAHULUAN

Pendidikan merupakan kebutuhan yang sangat penting bagi setiap individu, dan salah satu tujuan negara yang tercantum dalam Pembukaan UUD yaitu “Mencerdaskan Kehidupan Bangsa”. Pendidikan di Indonesia terbagi menjadi dua bagian yaitu pendidikan formal dan pendidikan informal. Kumon merupakan salah satu sarana pendidikan informal, yaitu tempat bimbingan belajar Matematika dan Bahasa Inggris yang memiliki metode pembelajaran tersendiri berdasarkan kemampuan individu atau berdasarkan level. Dengan metode Kumon banyak siswa yang belajar dengan level di atas tingkatan kelasnya sehingga siswa saat di sekolah dengan mudah mengikuti materi yang diberikan oleh guru karena telah melampaui level kelasnya, namun ada juga beberapa siswa yang belajar dengan di bawah tingkatan kelasnya atau setara tingkatan kelasnya. Ada beberapa faktor yang mempengaruhi tingkatan level pada siswa yaitu seperti: minat belajar, motivasi belajar, kebiasaan belajar dan jenis kelamin siswa.

“*Data Mining* adalah analisis otomatis dari data yang berjumlah besar atau kompleks dengan tujuan untuk menemukan pola atau kecenderungan yang penting yang biasanya tidak disadari keberadaannya.”[1]

Naïve Bayes Classifier merupakan salah satu metode data mining dalam pengklasifikasian yang terbukti memiliki akurasi dan kecepatan yang tinggi saat di aplikasikan kedalam *data base* dengan data

yang besar [2]. *Naïve Bayes* merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman dimasa sebelumnya. Dan dikombinasikan dengan *Naïve* dimana antar atribut atau variabelnya saling bebas dan tidak berkaitan satu sama lain [3].

Pada penelitian ini akan dilakukan analisis untuk memprediksi tingkatan level kumon pada siswa agar para pengajar dan orang tua dapat memberikan pembelajaran yang tepat agar siswa dapat belajar dengan di atas tingkatan kelasnya. Dalam menentukan tingkatan level pada siswa biasanya akan terlihat pada siswa saat mengikuti kelas kumon 6 bulan sampai dengan 1 tahun, namun dengan menggunakan metode *Naïve Bayes Classifier* kita dapat memprediksi tingkatan level kumon dengan mudah dan cepat.

Penelitian ini melakukan studi pendahuluan terhadap penelitian yang membahas “klasifikasi peminatan siswa SMA menggunakan Metode *Naïve Bayes*” [3], dengan tingkat akurasi 99.47 % dalam mengklasifikasikan status peminatan siswa SMA. Dan pada penelitian yang berjudul “*Naïve Bayes Classification* untuk penentuan status penduduk miskin dengan tingkat akurasi sebesar 70% [4].

Berdasarkan uraian di atas penelitian ini bertujuan untuk menentukan klasifikasi tingkatan level Kumon dengan menggunakan metode *Naïve Bayes Classifier* dan untuk mengetahui tingkat kepercayaan hasil klasifikasi tingkatan level Kumon dengan menggunakan metode *Naïve Bayes Classifier* dengan atribut minat belajar, motivasi belajar, kebiasaan belajar, dan jenis kelasmin. Dari sinilah penulis melakukan penelitian klasifikasi tingkatan level kumon dengan menggunakan metode *Naïve Bayes Classifier* di Kumon Otista Tangerang.

II. HASIL DAN PEMBAHASAN

II.1 Landasan Teori

Secara umum *Data Mining* terbagi atas 2 (dua) kata yaitu:

1. *Data* yaitu kumpulan fakta yang terekam atau sebuah entitas yang tidak memiliki arti dan selama ini terabaikan.
2. *Mining* yaitu proses penambangan. Sehingga *Data Mining* itu dapat diartikan sebagai proses penambangan data yang menghasilkan sebuah *output* (keluaran) berupa pengetahuan [1].

Naïve Bayesian Classifier merupakan salah satu algoritma pemecahan masalah yang termasuk kedalam metode klasifikasi pada *Data Mining*. *Naïve Bayes Classifier* mengadopsi ilmu statistika yaitu menggunakan teori kemungkinan (probabilitas) untuk menyelesaikan sebuah kasus *Supervised Learning*, artinya dalam himpunan data terdapat label, *class* atau target sebagai acuan atau gurunya. Persamaan dari teorema *Bayes* adalah [5] :

$$P(H | X) = \frac{P(X | H)P(H)}{P(X)}$$

Keterangan:

x : sampel data yang memiliki kelas yang tidak diketahui.

H : hipotesa bahwa x adalah data kelas.

P(H) : peluang dari hipotesa H.

P(X) : peluang dari data sampel yang diamati.

P(X | H) : peluang dari data sampel X bila diasumsikan bahwa hipotesa benar

RapidMiner adalah sebuah *software* untuk pengolahan *Data Mining*. *RapidMiner* adalah sebuah solusi untuk melakukan analisis terhadap *Data Mining*, *text mining* dan analisis prediksi [6].

Confussion matrix adalah suatu metode yang biasanya digunakan untuk melakukan perhitungan akurasi pada konsep *Data Mining* atau sistem pendukung keputusan [7] Berikut ini merupakan *Performance Confussion Matrix*:

$$Accurasi = \frac{TP+TN}{TP+TN+FP+FN} \times 100\%$$

$$Precision = \frac{TP}{TP+FP} \times 100\%$$

$$Recall = \frac{TP}{TP+FN} \times 100\%$$

Keterangan:

- TP : true positive
- TN : true negative
- FP : false positive
- FN : false negative

Validitas adalah suatu ukuran yang menunjukkan tingkat kesahihan suatu instrument". Suatu tes dikatakan valid apabila tes tersebut sesuai mengukur apa yang hendak diukur. Tes memiliki validitas yang tinggi jika hasilnya sesuai dengan kriteria, agar data yang terkumpul tidak menyimpang dari gambaran variabel yang digunakan [8] Uji validitas dapat dihitung dengan menggunakan korelasi *product moment* sebagai berikut:

$$r_{xy} = \frac{n \sum x_i y_i - (\sum x_i)(\sum y_i)}{\sqrt{(n \sum x_i^2 - (\sum x_i)^2)(n \sum y_i^2 - (\sum y_i)^2)}}$$

Keterangan:

- r_{xy} = Koefisien korelasi
- x = Skor item
- n = jumlah koresponden
- y = jumlah dari skor item
- Pernyataan dinyatakan valid jika nilai $r_{hitung} > r_{tabel}$.

Uji reliabilitas dilakukan untuk mengetahui seberapa jauh hasil pengukuran tetap konsisten apabila dilakukan pengukuran dua kali atau lebih terhadap gejala yang sama dengan menggunakan alat pengukur sama [9] Untuk mengetahui suatu alat ukur itu reliabel dapat diuji dengan menggunakan rumus *Cronbach Alpha* sebagai berikut:

$$r_i = \left(\frac{k}{k-1} \right) \left(1 - \frac{\sum \sigma_b^2}{\sigma_t^2} \right)$$

Keterangan:

- r_i = Reliabilitas Instrumen
- k = Banyaknya butir pertanyaan atau banyaknya soal
- $\sum \sigma_b^2$ = Jumlah varians butir
- σ_t^2 = Varians total

Tujuan kategorisasi adalah untuk menempatkan individu ke dalam kelompok

terpisah secara berjenjang menurut suatu kontinum berdasar atribut yang diukur [10] Berikut merupakan rumus yang digunakan dalam kategorisasi:

$$Mean empirik (\mu) = \frac{1}{n} \sum_{i=1}^n x_i$$

$$Standar deviasi populasi (\sigma) = \sqrt{\frac{\sum_{i=1}^n (x_i - \mu)^2}{n}}$$

Berdasarkan perhitungan di atas maka setiap responden akan di kelompokkan ke dalam tiga kategori [11] yaitu:

- Rendah = $x < \mu - 1 \sigma$
- Sedang = $\mu - 1 \sigma \leq x < \mu + 1 \sigma$
- Tinggi = $\mu + 1 \sigma \leq x$

II.2 Metodologi Penelitian

Penelitian ini merupakan penelitian deskriptif kuantitatif karena untuk menentukan klasifikasi tingkatan level kumon, variabel-variabel yang digunakan saling bebas dan tidak ada keterkaitan dengan variabel lainnya, dan data yang diperoleh dari sampel populasi berupa angka-angka dan dianalisis sesuai dengan metode statistik yang digunakan kemudian diinterpretasikan.

Penelitian dilakukan di Kumon Otista Tangerang dengan populasi siswa/siswi matematika sebanyak 80 sampel, dengan teknik pengambilan sampel yaitu *sampling jenuh*.

Data yang digunakan adalah data primer yang dilakukan melalui kuisioner atau angket untuk variabel minat belajar, motivasi belajar, dan kebiasaan belajar, yang kemudian akan dikategorikan menjadi tiga kategori untuk tiap-tiap variabel yaitu tinggi, rendah, dan sedang.

II.3 Pembahasan

Data training ini berisikan atribut-atribut yang mempengaruhi dalam menentukan tingkatan level kumon sebanyak 70 data training. Atribut pada penelitian ini adalah motivasi belajar, minat belajar, kebiasaan belajar dan jenis kelamin. Dan label atau kelas pada penelitian ini adalah di atas tingkatan kelas, di bawah tingkatan kelas, dan setara tingkatan kelas.

Data training ini yang digunakan untuk membuat sebuah model yang akan membentuk hasil dari *data testing* yang akan di uji.

Data testing yang digunakan pada penelitian ini diperoleh dari hasil penyebaran kuisioner pada siswa-siswi Matematika Kumon Otista sebanyak 10 *data testing* yang sudah diketahui tingkatan levelnya sehingga akan diteliti presentase keakuratan pada metode *Naïve Bayes Classifier* dengan menggunakan tabel *Confussion Matrix*.

II.4 Hasil Analisis Data

A. Probabilitas Label

1. $P(\text{tingkatan level : "di atas"})$

$$P(C_1) = 38/70 = 0,54$$

Terdapat 38 sampel pada tingkatan level di atas dari 70 *Data Training*.

2. $P(\text{tingkatan level : "di bawah"})$

$$P(C_2) = 14/70 = 0,20$$

Terdapat 14 sampel pada tingkatan level di bawah dari 70 *Data Training*.

3. $P(\text{tingkatan level : "setara"})$

$$P(C_3) = 18/70 = 0,26$$

Terdapat 18 sampel pada tingkatan level setara dari 70 *Data Training*.

B. Probabilitas Tiap Atribut

Tabel 1: Probabilitas Jenis Kelamin

P(JENISKELAMIN)	DI ATAS	DI BAWAH	SETARA
P	16/38	6/14	10/18
L	22/38	8/14	8/18

Tabel 2: Probabilitas Motivasi Belajar

P(MOTIVASI)	DI ATAS	DI BAWAH	SETARA
TINGGI	11/38	0/14	3/18
RENDAH	8/38	1/14	2/18
SEDANG	19/38	13/14	13/18

Tabel 3: Probabilitas Minat Belajar

P(MINAT)	DI ATAS	DI BAWAH	SETARA
TINGGI	11/38	0/14	2/18
RENDAH	8/38	4/14	3/18
SEDANG	19/38	10/14	13/18

Tabel 4: Probabilitas Kebiasaan Belajar

P(MOTIVASI)	DI ATAS	DI BAWAH	SETARA
TINGGI	11/38	0/14	2/18
RENDAH	8/38	2/14	4/18

SEDANG	19/38	12/14	12/18
--------	-------	-------	-------

C. Mengkalikan nilai atribut pada setiap Label.

Untuk sampel yang pertama jenis kelamin laki-laki, motivasi belajar tinggi, minat belajar sedang, dan kebiasaan belajar sedang.

$$\triangleright P(X1 | \text{tingkatan level : "di atas"})$$

$$= 22/38 \times 11/38 \times 19/38 \times 19/38 \\ = 0,042$$

$$\triangleright P(X1 | \text{tingkatan level : "di bawah"})$$

$$= 8/14 \times 0/14 \times 4/14 \times 12/14 \\ = 0$$

$$\triangleright P(X1 | \text{tingkatan level : "setara"})$$

$$= 8/18 \times 3/18 \times 13/18 \times 12/18 \\ = 0,036$$

D. Membandingkan nilai probabilitas setiap label

$$\triangleright P(X1 | \text{tingkatan level : "di atas"}) \times P(\text{tingkatan level : "di atas"})$$

$$= 0,042 \times 0,54 \\ = 0,022$$

$$\triangleright P(X1 | \text{tingkatan level : "di bawah"}) \times P(\text{tingkatan level : "di bawah"})$$

$$= 0 \times 0,2 \\ = 0$$

$$\triangleright P(X1 | \text{tingkatan level : "setara"}) \times P(\text{tingkatan level : "setara"})$$

$$= 0,036 \times 0,26 \\ = 0,009$$

Nilai maksimal pada perhitungan di atas adalah probabilitas untuk label di atas maka prediksi klasifikasi tingkatan level kumon pada sampel yang pertama adalah di atas tingkatan level kumon.

II.5 Uji Tabel *Confussion Matrix*

Pada perhitungan *Data testing* di atas maka hasil klasifikasi *Naïve Bayes Classifier* maka dapat di konfersi kedalam bentuk *Confussion Matrix* sebagai berikut:

Tabel 5: *Confussion Matrix*

		AKTUAL			TOTAL
		Di atas	Di bawah	Setara	
PREDIKSI	Di atas	3	1	0	4
	Di bawah	0	1	0	1
	Setara	0	2	3	5
TOTAL		3	4	3	10

Berdasarkan tabel 5 maka dapat di hitung *performance* dari metode *Naïve Bayes Classifier* sebagai berikut:

➤ *Accurasi*

$$\begin{aligned}
 &= \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \\
 &= \frac{3+(1+3)}{3+(1+3)+(1+0)+(2+0+0+0)} \times 100\% \\
 &= \frac{7}{10} \times 100\% \\
 &= 70\%
 \end{aligned}$$

➤ *Precision (di atas)*

$$\begin{aligned}
 &= \frac{TP}{TP+FP} \times 100\% \\
 &= \frac{3}{3+1} \times 100\% \\
 &= \frac{3}{4} \times 100\% \\
 &= 75\%
 \end{aligned}$$

➤ *Precision (di bawah)*

$$\begin{aligned}
 &= \frac{TP}{TP+FP} \times 100\% \\
 &= \frac{1}{1+0} \times 100\% \\
 &= 100\%
 \end{aligned}$$

➤ *Precision (setara)*

$$\begin{aligned}
 &= \frac{TP}{TP+FP} \times 100\% \\
 &= \frac{3}{3+2} \times 100\% \\
 &= \frac{3}{5} \times 100\% \\
 &= 60\%
 \end{aligned}$$

➤ *Recall (di atas)*

$$\begin{aligned}
 &= \frac{TP}{TP+FN} \times 100\% \\
 &= \frac{3}{3+0} \times 100\% \\
 &= 100\%
 \end{aligned}$$

➤ *Recall (di bawah)*

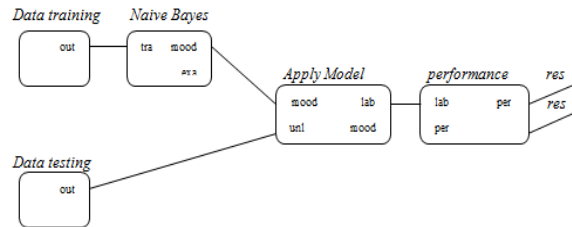
$$\begin{aligned}
 &= \frac{TP}{TP+FN} \times 100\% \\
 &= \frac{1}{1+(1+2)} \times 100\% \\
 &= \frac{1}{4} \times 100\% \\
 &= 25\%
 \end{aligned}$$

➤ *Recall (setara)*

$$\begin{aligned}
 &= \frac{TP}{TP+FN} \times 100\% \\
 &= \frac{3}{3+0} \times 100\% \\
 &= 100\%
 \end{aligned}$$

II.6 Uji *Shoftware RapidMiner*

Berikut ini merupakan *design* yang digunakan pada *software RapidMiner*:



Gambar 1: Design Performance

Dari perhitungan klasifikasi menggunakan *software RapidMiner* dengan menggunakan metode *Naïve Bayes Classifier* akan menampilkan probabilitas klasifikasi tingkatan level kumon hasil dari *data training* dan *data testing*. Dan untuk mengetahui tingkat *accuracy*, *class precision*, dan *recall* dapat dilihat pada gambar *Performance Vector*.

Tabel 6: Performance Vektor

Accuracy : 70%				
	True Di bawah	True Jantan	True Setara	Class Precision
Pred. Di atas	3	1	0	75%
Pred. Di bawah	0	1	0	100%
Pred. Setara	0	2	3	60%
Class Recall	100%	25%	100%	

Berdasarkan tabel 6 dapat kita ketahui tingkat *accuracy* sebesar 70%, *class recall* di atas sebesar 100%, *class recall* di bawah sebesar 25%, *class recall* setara sebesar 100%, *class precision* di atas sebesar 75%, *class precision* di bawah sebesar 100%, dan *class precision* setara sebesar 60%.

III.KESIMPULAN

➤ Untuk menentukan prediksi klasifikasi tingkatan level kumon dapat ditentukan dengan membandingkan nilai probabilitas setiap label atau kelas

dengan melihat nilai maksimum sebagai label atau kelas keputusan.

- Besar tingkat kepercayaan hasil klasifikasi tingkatan level Kumon dengan menggunakan metode Naïve Bayes Classifier yaitu nilai accuracy sebesar 70%, class recall di atas sebesar 100%, class recall di bawah sebesar 25%, class recall setara sebesar 100%, class precision di atas sebesar 75%, class precision di bawah sebesar 100%, dan class precision setara sebesar 60% sehingga dapat dinyatakan bahwa Algoritma Naïve Bayes Classifier dapat digunakan sebagai masukan untuk pengambil keputusan.
- Untuk penelitian berikutnya diharapkan menggunakan atribut-atribut yang lain dan menggunakan *data training* yang lebih banyak, agar menghasilkan nilai akurasi yang lebih tinggi.

Kualitatif, dan R&D. Bandung: Alfabeta, CV, 2010.

- [10] Azwar, *Realibilitas dan Validitas*. Yogyakarta: Pustaka Pelajar, 2012.
- [11] A. Saifudin, *Penyusunan Skala Psikologi*. Yogyakarta: Pustaka Pelajar, 2007.

REFERENSI

- [1] D. Nofriansyah and G. W. Nurcahyo, *Algoritma Data Mining Dan Pengujian*. In *Definisi dan Konsep Data Mining*. Yogyakarta: CV Budi Utama, 2015.
- [2] E. T. Kusriani, *Algoritma Data Mining*. Yogyakarta: Andi Offset, 2009.
- [3] H. Noparin, "Klasifikasi Peminatan Siswa SMA Menggunakan Metode Naive Bayes," *Syst. Inf. Syst. Informatics J.*, vol. 2, no. 1, pp. 25–32, 2016, doi: 10.29080/systemic.v2i1.104.
- [4] M. Rasyida, "Naïve Bayes Classification untuk Penentuan Status Penduduk Miskin," *J. Inform. Kaputama(JIK)*, vol. 4, no. 2, 2020.
- [5] D. Nofriansyah and G. W. Nurcahyo, *Algoritma Data Mining Dan Pengujian*. In *Algoritma Naive Bayesian classifier*. Yogyakarta: Deepublish, 2012.
- [6] A. Purwanto *et al.*, "Perbandingan Minat Siswa Smu Pada Metode Klasifikasi Menggunakan 5 Algoritma," vol. 2, no. 1, pp. 43–47, 2018.
- [7] P. E., *Konsep dan Aplikasi Menggunakan Matlab*. In *Andi, Data Mining*. Yogyakarta: Airlangga, 2012.
- [8] S. Arikunto, *Prosedur Penelitian Suatu Pendekatan Praktik*. Jakarta: Rineka Cipta, 2010.
- [9] Sugiyono, *Metode Penelitian Kuantitatif,*