

Algoritma *Naive Bayes* untuk Mengklasifikasikan Kepribadian Siswa SMP Berdasarkan Tipologi *Hippocrates-Galenus*

¹Meilana, ²Yuli Astuti*, ³Irma Rofni Wulandari, ⁴Istri Sulistyowati, ⁵Brigida Arie M.

¹Informatika, Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta

²Manajemen Informatika, Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta

³Sistem Informasi, Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta

⁴Teknik Informatika, Fakultas Ilmu Komputer, Universitas Widya Dharma Klaten

⁵Informatika, Fakultas Sains dan Komputer, Universitas Kristen Immanuel

^{1,2,3}Jl Ringroad Utara, Condongcatur, Depok, Sleman, Yogyakarta, Indonesia 55283

⁴Jl.Ki Hajar Dewantara, klaten Utara, Indonesia

⁵Jl.Solo Km 11, Yogyakarta

*Email : yuli@amikom.ac.id

(received: 16 Maret 2021, revised: 31 Maret 2021, accepted: 20 April 2021)

Abstrak

Perkembangan teknologi informasi berbasis komputer rentan memberikan dampak yang negatif terhadap siswa di usia remaja, secara khusus pada jenjang SMP yaitu usia 12 hingga 15 tahun. Usia tersebut merupakan masa peralihan usia anak-anak menuju remaja. Guru bimbingan konseling (BK) menjadi wadah untuk membenahi serta mendidik para siswa yang mempunyai permasalahan dari segi psikologi, karakteristik, ataupun hal pendukung lainnya. Namun pada prakteknya saat kegiatan konseling siswa akan merasa malu mengungkapkan permasalahan yang dihadapi dan cenderung menutupinya sehingga BK butuh banyak waktu untuk mengetahui permasalahan dan karakter siswa. Oleh karena itu diperlukan sistem untuk mengetahui karakteristik yang dimiliki oleh siswa, dengan tujuan agar BK dapat berkomunikasi dengan baik terhadap siswa yang memiliki permasalahan dan bisa mengetahui karakternya sehingga akan mempermudah dalam penanganannya. Penerapan algoritma *Naive Bayes Classifier* pada penelitian ini ditujukan untuk mengklasifikasikan data berdasarkan karakteristik yang sudah ditentukan sebelumnya, yakni *sanguin*, *koleris*, *melankolis*, *plegmatis*. Metode ini digunakan untuk menentukan kelas tertinggi yang akan ditujukan pada karakteristik kepribadian tersebut. Atribut yang diperlukan pada penelitian ini adalah nama siswa, usia, jenis kelamin, asal sekolah serta jawaban soal test A, B, C, dan D. Pengujian terhadap sistem dilakukan sebanyak 6 kali pengujian, dimana pengujian pertama mendapatkan nilai akurasi sebesar 68,57%, pengujian kedua sebesar 74%, pengujian ketiga sebesar 77,78%, pengujian keempat sebesar 81,18%, pengujian kelima sebesar 85,88% serta pengujian keenam sebesar 83,53%. Nilai akurasi tertinggi dihasilkan dari pengujian kelima sebesar 85,88% pengujian tersebut menggunakan *Confusion Matrix*, dalam hal ini Algoritma *Naive Bayes Classifier* mampu mengklasifikasikan dengan baik dan memperoleh hasil akurasi yang cukup tinggi.

Kata kunci : data mining, klasifikasi, *naive bayes classifier*, karakteristik kepribadian

Abstract

The development of computer-based information technology is prone to hurt teenage students, especially at the junior high school level, namely 12 to 15 years of age. This age is a period of transition from the ages of children to adolescents. The counseling guidance teacher (BK) becomes a place to fix and educate students who have psychological problems or other supporting matters. However, in practice, during counseling activities, students will reveal problems that arise and tend to cover them up so that counseling takes a lot of time to find out the problems and character of

<http://sistemasi.ftik.unisi.ac.id>

students. Therefore we need a system to see the contents of the students, with the aim that BK can communicate well with students who have problems and can see their character so that it will make it easier to handle. The application of the Naïve Bayes Classifier algorithm in this study is to classify data based on what has been predetermined, namely sanguine, choleric, melancholy, phlegmatic. This method is used to determine the highest class that will be aimed at that personality. The attributes required in this study are the name of the student, age, gender, school origin, and answers to test questions A, B, C, and D. Testing of the system was carried out 6 times, where the first test got an accuracy value of 68.57%, the second test was 74%, the third test was 77.78%, the fourth test was 81.18%, the fifth test was 85.88% and the sixth test was 83.53%. The highest value generated from this test is 85.88%. The test uses the Confusion Matrix, in this case, the Naïve Bayes Classifier Algorithm can classify well and get high enough results.

Keywords : data mining, classification, naïve bayes classifier, characteristics personality

1 Pendahuluan

Teknologi informasi berbasis komputer terus semakin berkembang, dimana berbagai aspek akan lebih dipermudah dengan adanya teknologi tersebut, tanpa terkecuali dalam dunia pendidikan[1]. Pendidikan merupakan media pembelajaran untuk menciptakan karakter, jati diri, serta norma-norma pada lingkungan sekolah maupun bermasyarakat. Sekolah Menengah Pertama (SMP) rata-rata berusia 12 sampai 15 tahun, dimana pada usia-usia tersebut merupakan masa peralihan dari anak-anak menuju remaja. Rasa ingin tahu terhadap dunia luar yang terkadang berkonteks negatif sering terjadi, pada hal ini dibutuhkan guru Bimbingan Konseling (BK) yang akan membantu menyelesaikan permasalahan tersebut[2][3].

Guru BK seperti menjadi wadah untuk membenahi serta mendidik para siswa yang mempunyai permasalahan yang menyimpang dalam hal psikologi, karakteristik, ataupun hal pendukung lainnya. Tetapi akan lebih mudah jika dalam membimbing dan membantu menyelesaikan permasalahan siswa, BK juga mengetahui karakteristik seperti apakah yang dimiliki para siswanya[4][5]. Menurut[6] bimbingan konseling juga bisa dilakukan dengan bantuan system untuk mengurangi rasa takut terhadap BK dan akan menghemat waktu karena tidak seimbang jumlah siswa dengan guru BK pada setiap sekolah. Karakteristik kepribadian bisa dikaji dan dipelajari melalui teori tentang kepribadian manusia yang sudah dikemukakan oleh tokoh-tokoh terdahulu. Salah satunya yaitu teori *hippocrates-galenus* yang mengemukakan bahwa kepribadian manusia dibedakan menjadi empat yaitu melankolis, sanguinis, plegmatis dan koleris[7].

Menurut[8] algoritma *Naïve Bayes Classifier* merupakan algoritma yang didalamnya terdapat teknik klasifikasi, dimana pengklasifikasian tersebut menggunakan metode probabilitas serta statistik, selain itu juga mudah diimplementasikan dan memberikan hasil yang akurat.

Algoritma *Naïve Bayes Classifier* dikategorikan sebagai salah satu klasifikasi dengan efisiensi dan akurasi yang baik, kelebihan pada algoritma ini adalah menangani data kuantitatif dan data diskrit, hanya memerlukan sejumlah data yang relatif kecil. Adapun menurut[9] pada algoritma ini terdapat kelemahan dimana sebuah probabilitas tidak bisa mengukur pada saat dimana ditetapkan seberapa besar tingkat keakuratan sebuah prediksi. Hal lain yang menunjukkan kelemahan pada algoritma *Naïve Bayes Classifier* terdapat pada sebuah seleksi atribut sehingga dapat mempengaruhi nilai akurasi.

Banyak yang menggunakan algoritma *Naïve Bayes Classifier* karena tingkat akurasi yang tinggi, salah satunya dalam bidang maskapai penerbangan. Adapun penelitian tersebut dengan tujuan untuk menguji algoritmanya itu sendiri, dimana hasil akhir akan menentukan klasifikasi yang mempunyai nilai terbesar dan ingin mendapatkan hasil yang akurat dari setiap permasalahan yang diuraikan dan diteliti dengan algoritma *Naïve Bayes Classifier*[10]. Sedangkan pada penelitian ini bertujuan untuk menguji tingkat akurasi algoritma Naïve Bayes pada berdasarkan klasifikasi karakter siswa.

2 Tinjauan Pustaka

Penelitian yang berjudul “*Analisis Data Mining Untuk Pemetaan Mahasiswa Yang Membutuhkan Bimbingan dan Konseling Menggunakan Algoritma Naïve Bayes Classifier*”, menjabarkan keberhasilan mahasiswa dalam menjalani proses pendidikan di perguruan tinggi. Keberhasilan tersebut diukur melalui kinerja akademik dari mahasiswa itu sendiri. Namun terlepas dari keberhasilan itu terdapat pula beberapa permasalahan yang dihadapi mahasiswa dari lingkungan luar maupun di dalam lingkungan pendidikan. Maka dari itu institusi mendorong peningkatan kinerja akademik untuk memberikan pelayanan berupa Bimbingan dan Konseling (BK). Algoritma yang digunakan pada penelitian ini yaitu *Naïve Bayes Classifier* dengan data mining untuk mengklasifikasikan mahasiswa yang memiliki kinerja akademik baik maupun buruk. Guna menunjang keberhasilan penelitian ini, dibutuhkan sebanyak 507 recordset data akademik mahasiswa tahun masuk 2008, program S1 Fakultas Teknik Universitas Negeri Padang. Algoritma tersebut berhasil berhasil mengklasifikasi data mahasiswa yang membutuhkan bimbingan konseling yang telah diadakan oleh pusat pelayanan akademik. Hasil akurasi yang diperoleh cukup besar mewakili keseluruhan program studi yaitu 81% [11].

Penelitian lainnya dengan judul “*Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naïve Bayes Classifier (Studi Kasus : Fasilkom Unilak)*” mengurai tentang kinerja akademik mahasiswa dalam menyelesaikan masa studi terhitung dari para mahasiswa berada pada tahun ke 3 atau tepatnya berada di semester 6. Evaluasi pada penelitian ini dilakukan dengan membagi data menjadi dua yaitu data latih dan data uji. Data latih diambil dari data lulusan mahasiswa semester 6 sebanyak 50 data, sedangkan data uji diambil dari mahasiswa semester 6 yang belum menyelesaikan masa studinya sebanyak 61 data. Nilai keakuratan terbaik pada penelitian ini terhitung 76%, dengan perolehan hasil 48 mahasiswa pada tahun ke 3 atau semester 6 lulus tepat waktu dan 13 mahasiswa lulus tidak tepat waktu. Atribut yang digunakan pada penelitian ini diantaranya adalah *jenis kelamin, program studi, asal sekolah, daerah asal, tempat tanggal lahir, indeks prestasi semester 1 sampai dengan 5, indeks prestasi kumulatif dan target kelulusan* [12].

Serupa dengan penelitian sebelumnya, penelitian lainnya yakni “*Penerapan Metode Naïve Bayes untuk Klasifikasi SMS Spam Menggunakan Java Programming*”. Pada penelitian ini penulis menjabarkan tentang klasifikasi *Short Message Service (SMS) spam*. SMS itu merupakan layanan komunikasi yang digunakan seseorang untuk mengirim atau menerima pesan singkat berupa text pada ponsel yang dimiliki pengguna. Pada penelitian ini metode yang digunakan yaitu algoritma *Naïve Bayes* secara umum. Data latih yang digunakan adalah SMS spam dan SMS bukan spam, aplikasi ini mempunyai morfologi tertentu dalam proses pengkategorian. Tahap pemrosesan pada aplikasi ini diantaranya yaitu *preprocessing* berupa *case folding* dan *parsing, transformation* berupa penghapusan *stopword removal* dan *stemming*. Akurasi yang diperoleh pada penelitian ini sendiri yaitu rata-rata *precision* sebesar 24%, *recall* 88% dan *confusion matrix* sebesar 62% [13][14].

2.1 Dasar Teori

Kepribadian adalah sejarah sosial yang terdapat pada diri manusia yang bersifat alami serta dapat membedakan dirinya dari sesuatu yang mengelilingi kehidupannya. Dapat dikatakan kepribadian yang sehat apabila tercapainya taraf diferensiasi dan perkembangan yang optimal, selain itu tingkah laku merupakan faktor utama yang bisa memaknai sebuah kepribadian manusia. Dalam dunia psikologi menyebutkan bahwa kepribadian adalah sebagai satu kesatuan antara jiwa dan raga setiap manusia, hal itu juga membentuk sebuah kesadaran dan eksistensi yang dilakukan oleh setiap individu dan menekan pada pola perilaku yang autentik dan sangat unik dalam diri manusia [13][15].

2.2 Sejarah Sejarah Tipologi Hippocrates-Galenus

Menurut [16] tipologi merupakan sebuah pengetahuan yang menggolongkan setiap manusia menjadi tipe yang berbeda-beda berdasarkan faktor-faktor diantaranya fisik, psikis, pengaruh nilai-nilai budaya yang dominan dan lain sebagainya. Tipologi sendiri terbagi berbagai macam salah satunya yaitu tipologi konstitusi, beberapa ahli telah mengembangkan tipologi ini di antaranya adalah Hippocrates dan Galenus .

Tipologi ini dikembangkan oleh Galenus dimana berdasarkan pemikiran Hippocrates (460-370 SM) dipengaruhi oleh pandang Empedocles (490-435 SM) dimana tokoh tersebut mengemukakan bahwa pada alam semesta memiliki empat unsur dasar yaitu tanah, air, udara dan api. Berdasarkan pendapat yang disampaikan oleh Empedocles setelah itu Hippocrates menyatakan bahwa setiap manusia mempunyai 4 macam cairan yang memiliki sifat sama seperti unsur pada alam, sifat-sifat tersebut di antaranya yaitu :

- Sifat kering yang didefinisikan sebagai empedu kuning, dimiliki oleh *chloe* serta berwarna kuning.
- Sifat basah yang didefinisikan sebagai empedu hitam, dimiliki oleh *Melancolie* serta berwarna hitam
- Sifat dingin yang didefinisikan sebagai lendir, dimiliki oleh *flegma* serta berwarna putih
- Sifat panas yang didefinisikan sebagai darah, dimiliki oleh *sanguine* serta berwarna merah.

Menurut Hippocrates, keempat cairan tersebut yang dimiliki oleh individu tidak akan sama dengan milik individu lainnya. Setiap individu didominasi oleh salah satu cairan tersebut sehingga menyebabkan timbulnya ciri khas pada setiap orang. Galenus (129-199 SM) menyamakan pendapatnya dengan Hippocrates bahwasanya dalam diri seseorang terdapat 4 macam cairan tersebut. Namun hanya salah satu cairan yang mendominasi diri seseorang sehingga timbul ciri khas tertentu. Adapun sifat kejiwaan seseorang yang memiliki ciri khas tertentu akibat dominannya salah satu cairan menurut Galenus disebut temperamen.

2.3 Data Mining

Menurut[17] *data mining* merupakan proses mencari informasi menarik yang terdapat didalam data menggunakan metode-metode tertentu. Teknik dalam data mining sangat bervariasi, serta pemilihan algoritma yang tepat sangat dibutuhkan yang kemudian bergantung pada tujuan serta proses *Knowledge Discovery in Database (KDD)* secara keseluruhan.

2.4 Algoritma Naïve Bayes Classifier

Menurut[10] klasifikasi algoritma *Naïve Bayes Classifier* merupakan salah satu algoritma yang diasumsikan bahwa ada ataupun tidak ada ciri-ciri tertentu dari sebuah kelas maka tidak ada hubungannya dengan ciri-ciri dari kelas yang lainnya. Adapun bentuk umum dari teorema Bayes tertera pada persamaan 1.

$$P(Y|X) = \frac{P(X|Y).P(Y)}{P(X)} \quad (1)$$

Dimana :

X : data dengan *class* yang belum diketahui

Y : hipotesis dari data X, merupakan sebuah *class* yang spesifik

P(X) : probabilitas awal (priori) bukti X terjadi tanpa mengandung *hipotesis* yang lain

P(X|Y) : probabilitas sebuah bukti X terjadi akan mempengaruhi *hipotesis* Y

P(Y) : probabilitas awal (priori) *hipotesis* Y terjadi tanpa mengandung bukti apapun

P(Y|X) : probabilitas akhir bersyarat (conditional probability) suatu *hipotesis* Y terjadi jika diberikan bukti *evidence* X terjadi

Formulasi *Naïve Bayes Classifier* untuk klasifikasi adalah sebagai berikut terdapat pada persamaan 2.

$$P(Y|X) = \frac{P(Y) \prod_{i=1}^q P(X_i | Y)}{P(X)} \quad (2)$$

Dimana :

P(Y) : probabilitas awal *class* Y

P(X|Y) : probabilitas data dengan vektor X pada *class* Y

$\prod_{i=1}^q P(X_i | Y)$: probabilitas independen *class* Y dari semua fitur dalam vektor X

Distribusi *Gaussian* dipilih untuk merepresentasikan sebuah peluang *class* yang mempunyai sebuah syarat untuk atribut kontinyu. Distribusi ini mempunyai karakteristik dua parameter yaitu *mean* μ dan *varian* σ^2 untuk setiap *class* Y, sedangkan peluang bersyarat untuk atribut X terdapat pada persamaan 3.

$$P(= x_i | Y = y_i) = \frac{1}{\sqrt{2\pi\sigma_{ij}}} \exp \frac{(x_i - \mu_{ij})^2}{2\sigma_{ij}^2} \quad (3)$$

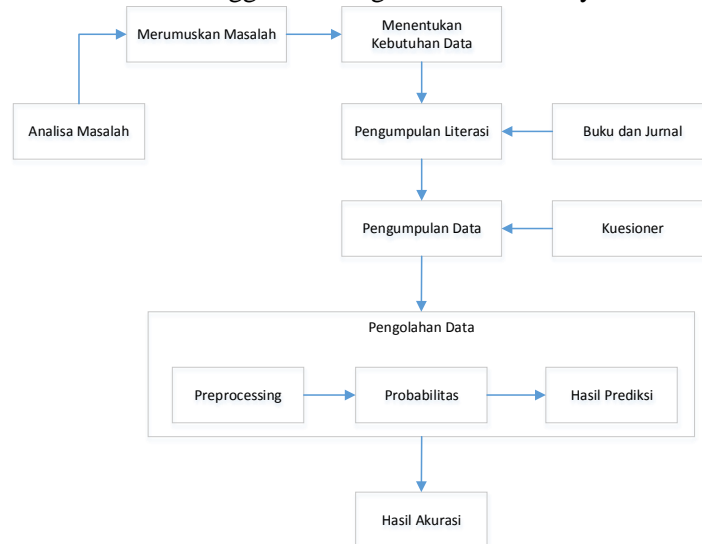
Dimana :

μ_{ij} : mean sample X_i (x) dari keseluruhan data latih

$2\sigma_{ij}^2$: sample (s^2) dari data latih

3 Metode Penelitian

Metode pada penelitian ini terdapat dua alur, pada Gambar 1 adalah diagram alir proses penelitian dan Gambar 2 adalah flowchart klasifikasi menggunakan algoritma Naïve Bayes.



Gambar 1 Diagram Alir Proses Penelitian



Gambar 2 Flowchart Klasifikasi menggunakan Naïve Bayes

<http://sistemasi.ftik.unisi.ac.id>

Pada Gambar 1 dan Gambar 2 menjelaskan tahapan-tahapan yang dilakukan pada penelitian ini. Tahap ini dimulai dari analisa masalah kemudian merumuskan permasalahan tersebut. Pada penelitian ini dataset yang digunakan berupa data karakteristik kepribadian peserta didik, dengan kebutuhan data nama, usia, jenis kelamin, asal sekolah, kelas, nomor handphone dan soal tes. Pada soal tes terdapat 40 soal tes yang terdapat didalam kuesioner online, yang masing-masing nomor memiliki pilihan jawaban yaitu A, B, C, dan D. Adapun data yang diperoleh dari pengisian kuesioner online yang dikumpulkan oleh siswa-siswi SMP N kelas 7 dan kelas 8 yang masing-masing terdiri dari 4 kelas yaitu 7A, 7B, 7C, 7D serta 8A, 8B, 8C, 8D pada bulan September 2020. Data diperoleh melalui kuesioner yang telah diisi dan dikumpulkan oleh siswa-siswi kelas 7 dan 8 SMP Negeri 1 Depok secara online. Kuesioner diklasifikasikan kedalam 4 kelas tipe kepribadian menurut tipologi *Hippocrates-Galenus* yaitu Sanguin, Koleris, Melankolis serta Plegmatis. Tahapan dalam pengolahan data adalah sebagai berikut.

1. Studi literatur digunakan sebagai acuan dalam melakukan proses perhitungan nilai-nilai probabilitas pada masing-masing kelas terhadap data latih sesuai dengan kaidah perhitungan algoritma *Naïve Bayes*. Selanjutnya nilai probabilitas akhir akan menentukan kategori kelas sesuai data yang sudah diuji.
2. Setelah data yang diperoleh dari kuesioner terkumpul langkah selanjutnya yaitu melakukan *preprocessing* data, yang awalnya hanya data mentah kemudian diproses dan dikelompokan berdasarkan tiap-tiap kelas yaitu Sanguin, Koleris, Melankolis, Plegmatis.
3. Pengolahan data pada tahap ketiga dilakukan setelah hasil dari *preprocessing* untuk menentukan hasil dari data yang bertipe numerik, diantaranya yaitu : menentukan standar deviasi, probabilitas, serta menentukan hasil akhir berdasarkan kaidah perhitungan algoritma *Naïve Bayes*.

Data yang didapatkan dari kuesioner kemudian diolah, tahap pengolahan data ini terdiri dari tiga tahapan yaitu *preprocessing* data kemudian dilanjutkan probabilitas menggunakan algoritma *naïve bayes* yang nantinya menghasilkan nilai prediksi, dari hasil nilai prediksi kemudian diolah lagi menggunakan *confusion matrix* untuk mendapatkan akurasi. Pada data uji menggunakan 10 data sampel, data sampel ini diperoleh dari kuesioner yang sudah diisi oleh para siswa-siswi. Dimana pada data uji didalamnya terdapat atribut Nama, Usia, Jenis Kelamin, Asal Sekolah, Jawaban Tes serta Kelas kepribadian.

4 Hasil dan Pembahasan

Data yang diperoleh dari pengisian kuesioner online yang dikumpulkan oleh siswa-siswi SMP N kelas 7 dan kelas 8 kemudian dilakukan *preprocessing* data dan sudah dikelompokan masing-masing menjadi 4 kelas karakteristik kepribadian yaitu Sanguin, Koleris, Melankolis, Plegmatis. Terdapat 100 record data yang dibagi menjadi 25 record data pada kelas *Sanguin*, 25 record data pada kelas *Koleris*, 25 record data pada kelas *Melankolis* dan 25 record data pada kelas *Plegmatis*. Sedangkan pada data uji menggunakan 10 record data yang terdiri dari kelas Sanguin, Koleris, Melankolis, Plegmatis. Dari data latih tersebut, kemudian dihitung nilai probabilitas kelasnya. Nilai Probabilitas yang akan dihitung terbagi menjadi 4, dimana diantara kelas-kelas tersebut sudah dibagi menjadi 25 jumlah prediksi setiap kelas yaitu Sanguin 25, Koleris 25 Melankolis 25 dan Plegmatis 25. Kemudian dilakukan perhitungan probabilitas sesuai dengan persamaan 1.

$$\begin{aligned} P(\text{Sanguin}) &= \sum \text{Sanguin} / \text{jumlah total} = 25/100 = 0,25 \\ P(\text{Koleris}) &= \sum \text{Koleris} / \text{jumlah total} = 25/100 = 0,25 \\ P(\text{Melankolis}) &= \sum \text{Melankolis} / \text{jumlah total} = 25/100 = 0,2556 \\ P(\text{Plegmatis}) &= \sum \text{Plegmatis} / \text{jumlah total} = 25/100 = 0,25 \end{aligned}$$

Pada semua fitur (fitur jenis kelamin, asal sekolah, usia, dan jawaban tes) juga akan dihitung nilai probabilitasnya dengan hasil terlihat pada Tabel 2 dan semua jumlah fitur terlihat pada Tabel 1.

Tabel 1 Jumlah Nilai Fitur

	Sanguin	Koleris	Melankolis	Plegmatis
Σ Laki-laki (L)	10	13	6	17
Σ Perempuan (P)	15	12	19	8
Σ Swasta (S)	1	5	1	2
Σ Negeri (N)	24	20	24	23
Σ Usia	324	321	323	327
Σ Jawaban Tes A	389	232	162	226
Σ Jawaban Tes B	147	341	163	140
Σ Jawaban Tes C	232	263	423	251
Σ Jawaban Tes D	232	164	252	383

Kemudian dihitung probabilitasnya menggunakan persamaan 1 dengan hasil terdapat pada Tabel 2.

Tabel 2 Nilai Probabilitas Semua Fitur

	Sanguin	Koleris	Melankolis	Plegmatis
P (Laki-laki)	$10/25 = 0,4$	$13/25 = 0,52$	$6/25 = 0,24$	$17/25 = 0,68$
P (Perempuan)	$15/25 = 0,6$	$12/25 = 0,48$	$19/25 = 0,76$	$8/25 = 0,32$
P (Swasta)	$1/25 = 0,04$	$5/25 = 0,2$	$1/25 = 0,04$	$2/25 = 0,08$
P (Negeri)	$24/25 = 0,96$	$20/25 = 0,8$	$24/25 = 0,96$	$23/25 = 0,92$
P (Usia)	$324/25 = 12,96$	$321/25 = 12,84$	$323/25 = 12,92$	$327/25 = 13,08$
P (Jawaban A)	$389/25 = 15,56$	$232/25 = 9,28$	$162/25 = 6,48$	$226/25 = 9,04$
P (Jawaban B)	$147/25 = 5,88$	$341/25 = 13,64$	$163/25 = 6,52$	$140/25 = 5,6$
P (Jawaban C)	$232/25 = 9,28$	$263/25 = 10,52$	$423/25 = 16,93$	$251/25 = 10,04$
P (Jawaban D)	$232/25 = 9,28$	$164/25 = 6,56$	$252/25 = 10,08$	$383/25 = 9,96$

Selanjutnya dihitung nilai standar deviasinya pada semua fitur, hasil nilai standar deviasi semua fitur terdapat pada Tabel 3.

Tabel 3 Nilai Standar Deviasi Semua Fitur

	Sanguin	Koleris	Melankolis	Plegmatis
S (Usia)	$16,2/24 = 0,675$ $\sqrt{0,675} = 0,82159$	$11,36/24 = 0,473$ $\sqrt{0,473} = 0,68774$	$7,84/24 = 0,326$ $\sqrt{0,326} = 0,57096$	$7,84/24 = 0,326$ $\sqrt{0,326} = 0,57096$
S (Jawaban A)	$196,16/24 = 8,1733$ $\sqrt{8,1733} = 2,8589$	$301,04/24 = 12,5433$ $\sqrt{12,5433} = 3,5416$	$274,16/24 = 11,4233$ $\sqrt{11,4233} = 3,3798$	$352,96/24 = 14,7066$ $\sqrt{14,7066} = 3,8349$
S (Jawaban B)	$152,64/24 = 6,36$ $\sqrt{6,36} = 2,5219$	$273,76/24 = 11,4066$ $\sqrt{11,4066} = 3,3773$	$226,24/24 = 9,4266$ $\sqrt{9,4266} = 3,0702$	$148/24 = 6,1666$ $\sqrt{6,1666} = 2,4832$
S (Jawaban C)	$151,04/24 = 6,2933$ $\sqrt{6,2933} = 2,5086$	$116,24/24 = 4,8433$ $\sqrt{4,8433} = 2,2007$	$141,84/24 = 5,91$ $\sqrt{5,91} = 2,4310$	$174,2/24 = 7,2583$ $\sqrt{7,2583} = 2,6941$
S (Jawaban D)	$239,04/24 = 9,96$ $\sqrt{9,96} = 3,1559$	$276,16/24 = 11,5066$ $\sqrt{11,5066} = 3,3921$	$230,68/24 = 9,6116$ $\sqrt{9,6116} = 3,1002$	$135,44/24 = 5,6433$ $\sqrt{5,6433} = 2,3755$

Pada nilai standar deviasi perhitungan hanya dilakukan pada fitur Usia, Jawaban Tes A, Jawaban Tes B, Jawaban Tes C serta Jawaban Tes D, dikarenakan semua masing-masing data pada fitur tersebut mempunyai nilai. Sedangkan pada fitur Jenis Kelamin dan Asal sekolah ada yang tidak mempunyai nilai sehingga pada fitur tersebut tidak bisa dihitung keseluruhan nilai standar deviasi pada kelas-kelas Sanguin, Koleris, Melankolis maupun Plegmatis. Selanjutnya adalah proses perhitungan dengan data uji, sample data uji yang digunakan ada 10 record, data uji tersebut dapat dilihat pada Tabel 4.

Tabel 4 Data Uji

No	Nama	Usia	JK	Asal Sekolah	Jawaban Tes				Kelas
					A	B	C	D	
1	Ramzy Farhah M	14	L	Negeri	22	12	4	2	Sanguin
2	Irene Dewi Vania Christiany	12	P	Negeri	11	9	11	9	Koleris
3	Ellena Novitasari	12	P	Negeri	16	7	8	9	Sanguin
4	Zulfia Ali Djumaeni	13	P	Negeri	8	14	13	5	Koleris
5	Elizabeth Velin Girik Allo	12	P	Negeri	14	3	12	11	Sanguin
6	Aqiel Zidan Khoiron	13	L	Swasta	4	8	12	16	Plegmatis
7	Naufal Ihsan Ardiansyah	13	L	Negeri	7	13	8	12	Plegmatis
8	Shafira Imas Sofwatunnisa	13	P	Negeri	6	5	17	12	Melankolis
9	Nathanael Nandana	13	L	Negeri	13	16	10	1	Koleris
10	Nadhira Lulu Cahyani	13	P	Negeri	13	4	14	9	Melankolis

Pada perhitungan Fitur Usia, Jawaban tes A,B,C,D penulis mengambil sampel dari data uji pertama, yaitu dengan Nama Ramzy Farhah, Usia 14th, Jenis Kelamin L (laki-laki), Asal sekolah Negeri, Jawaban tes A 22, Jawaban tes B 12, Jawaban tes C 4, Jawaban tes D 2, Kelas Sanguin. Selanjutnya dari data uji yang ada pada Tabel 4 akan dilakukan perhitungan ke semua fitur menggunakan persamaan 3, dengan hasil terlihat pada Tabel 5.

Tabel 5 Perhitungan Fitur Usia Pada Data Uji

	Sanguin	Koleris	Melankolis	Plegmatis
Usia	0,140644596	0,028791275	0,00295428	0,013231895
Jawaban A	0,102331387	0,067376207	0,046913874	0,06959861
Jawaban B	0,099591812	0,116937031	0,109763383	0,093778665
Jawaban C	0,111874597	0,073271225	0,015046955	0,104769311
Jawaban D	0,096800971	0,108753442	0,090399562	0,010363686

4.1 Menghitung Nilai probabilitas Akhir pada masing-masing data uji

Setelah mengetahui nilai pada semua fitur, selanjutnya menghitung nilai probabilitas akhir pada masing-masing data uji.

a. Kelas Sanguin

$$\begin{aligned}
 P(X | \text{Sanguin}) &= P(\text{Sanguin}) * P(\text{Jenis Kelamin} = \text{Laki-Laki} | \text{Sanguin}) * P(\text{Usia} = 14 | \text{Sanguin}) * \\
 &P(\text{Asal Sekolah} = \text{Negeri} | \text{Sanguin}) * P(\text{Jawaban Tes A} = 22 | \text{Sanguin}) * \\
 &P(\text{Jawaban Tes B} = 12 | \text{Sanguin}) * P(\text{Jawaban Tes C} = 4 | \text{Sanguin}) * P(\text{Jawaban} \\
 &\text{Tes D} = 2 | \text{Sanguin}) \\
 &= 0,25 * 0,4 * 0,96 * 0,1406 * 0,1023 * 0,0996 * 0,1119 * 0,0968 = 0,0000014896
 \end{aligned}$$

b. Kelas Koleris

$$\begin{aligned}
 P(X | \text{Koleris}) &= P(\text{Koleris}) * P(\text{Jenis Kelamin} = \text{Laki-Laki} | \text{Koleris}) * P(\text{Usia} = 14 | \text{Koleris}) * \\
 &P(\text{Asal Sekolah} = \text{Negeri} | \text{Koleris}) * P(\text{Jawaban Tes A} = 22 | \text{Koleris}) * P(\text{Jawaban} \\
 &\text{Tes B} = 12 | \text{Koleris}) * P(\text{Jawaban Tes C} = 4 | \text{Koleris}) * P(\text{Jawaban Tes D} = 2 | \\
 &\text{Koleris}) \\
 &= 0,25 * 0,52 * 0,8 * 0,0288 * 0,0674 * 0,1169 * 0,0733 * 0,108 = 0,0000001868
 \end{aligned}$$

c. Kelas Melankolis

$$\begin{aligned}
 P(X | \text{Melankolis}) &= P(\text{Melankolis}) * P(\text{Jenis Kelamin} = \text{Laki-Laki} | \text{Melankolis}) * P(\text{Usia} = 14 | \\
 &\text{Melankolis}) * P(\text{Asal Sekolah} = \text{Negeri} | \text{Melankolis}) * P(\text{Jawaban Tes A} = 22 |
 \end{aligned}$$

$$\begin{aligned} & \text{Melankolis}) * P(\text{Jawaban Tes B} = 12 \mid \text{Melankolis}) * P(\text{Jawaban Tes C} = 4 \\ & \text{Melankolis}) * P(\text{Jawaban Tes D} = 2 \mid \text{Melankolis}) \\ & = 0,25 * 0,24 * 0,96 * 0,003 * 0,0469 * 0,1098 * 0,015 * 0,0904 = 0,0000000121 \end{aligned}$$

d. Kelas Plegmatis

$$\begin{aligned} P(X \mid \text{Plegmatis}) &= P(\text{Plegmatis}) * P(\text{Jenis Kelamin} = \text{Laki-Laki} \mid \text{Plegmatis}) * P(\text{Usia} = 14 \mid \\ & \text{Plegmatis}) * P(\text{Asal Sekolah} = \text{Negeri} \mid \text{Plegmatis}) * P(\text{Jawaban Tes A} = 22 \mid \\ & \text{Plegmatis}) * P(\text{Jawaban Tes B} = 12 \mid \text{Plegmatis}) * P(\text{Jawaban Tes C} = 4 \mid \\ & \text{Plegmatis}) * P(\text{Jawaban Tes D} = 2 \mid \text{Plegmatis}) \\ & = 0,25 * 0,68 * 0,92 * 0,0132 * 0,0696 * 0,0938 * 0,1048 * 0,015 = 0,0000000212 \end{aligned}$$

Maka dapat disimpulkan nilai probabilitas akhir terbesar ada pada kelas Sanguin, oleh sebab itu data uji bisa diklasifikasikan sebagai karakteristik Sanguin.

4.2 Pengujian

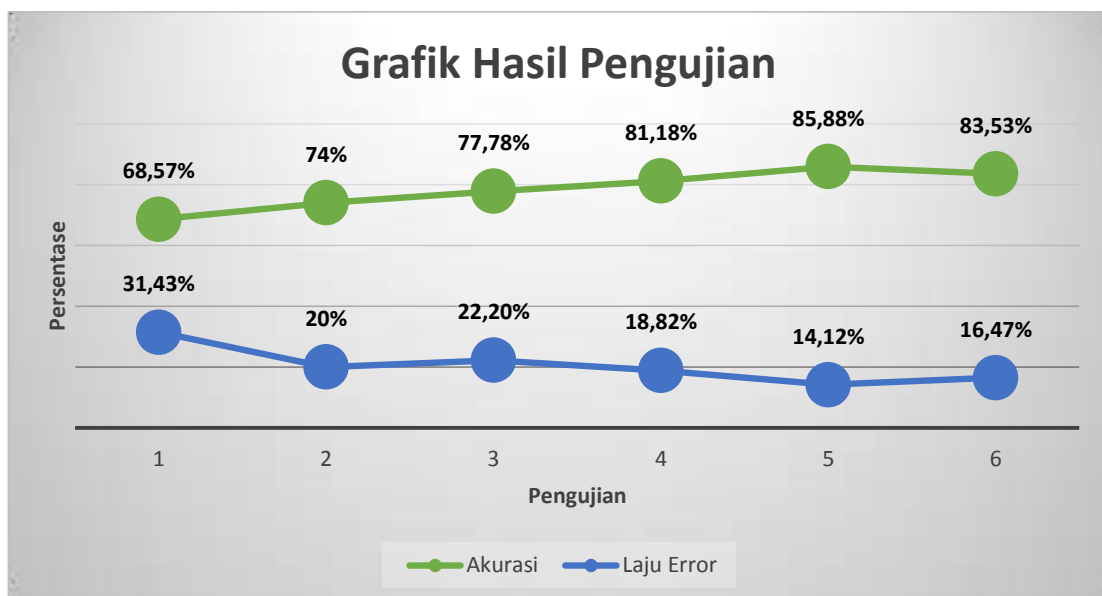
Pada penelitian ini dilakukan enam kali pengujian dengan data pengujian yang berbeda-beda jumlah recordnya. Hasil dari pengujian terdapat pada Tabel 6.

Tabel 6 Hasil Pengujian

Pengujian Ke	Data Pengujian	Data Latih				Data Uji				Akurasi	Laju Error
		S	K	M	P	S	K	M	P		
1	170	32	21	41	41	13	9	7	6	68,57%	31,43%
2	215	46	15	48	56	12	11	14	13	74%	20%
3	278	61	14	69	71	18	13	17	15	77,78%	22,2%
4	320	58	16	79	82	20	13	21	31	81,18%	18,82%
5	320	56	19	82	78	22	10	18	35	85,88%	14,12%
6	320	46	19	89	81	32	10	11	32	83,53%	16,47%

4.3 Hasil Pengujian Akurasi Algoritma Naïve Bayes Classifier

Pada pengujian pertama sampai keenam menunjukkan nilai akurasi yang berbeda, hal ini disebabkan jumlah komposisi data setiap kelas karakteristik kepribadian berbeda-beda. Oleh sebab itu semakin banyak data latih yang digunakan akan menambah variasi atau komposisi data sehingga nilai akurasi yang dihasilkan juga semakin baik. Dari enam pengujian, akurasi tertinggi didapat pada pengujian kelima yaitu dengan nilai akurasi sebesar 85,88% dan nilai laju error sebesar 14,12%. Data yang digunakan untuk pengujian kelima sebesar 320 record data, dimana 235 record data sebagai data latih dan 85 record data sebagai data uji. Adapun grafik hasil pengujian akurasi Algoritma Naïve Bayes Classifier dapat dilihat pada Gambar 3.



Gambar 3 Grafik Hasil Pengujian Akurasi Algoritma Naïve Bayes Classifier

5 Kesimpulan

Algoritma *Naïve Bayes Classifier* mampu mengklasifikasikan karakteristik kepribadian siswa dengan karakter kepribadian *sanguin, koleris, melankolis* serta *plegmatis*. Jumlah data latih pada proses klasifikasi akan sangat mempengaruhi nilai akurasi dan laju error, semakin banyak data latih yang digunakan maka nilai akurasi yang dihasilkan juga semakin baik. Nilai akurasi tertinggi sebesar 85,88% dan nilai laju error sebesar 14,12% pada pengujian kelima dari enam pengujian menggunakan metode *Confusion matrix*, sehingga dapat disimpulkan algoritma *Naïve Bayes Classifier* mampu mengklasifikasikan dengan baik dan memperoleh hasil akurasi yang cukup tinggi.

Referensi

- [1] M. Husaini, "Pemanfaatan Teknologi Informasi Dalam Bidang Pendidikan (E-Education)," *Mikrotik*, vol. 2, 2014.
- [2] H. Hermino, "Pendidikan Karakter Dalam Perspektif Psikologis Siswa Sekolah Menengah Pertama Diera Globalisasi dan Multikultural," *Peradaban*, vol. VIII, pp. 19–40, 2015.
- [3] Y. N. . Sari, "Perkembangan Kognitif dan Emosi Psikolog Masa Remaja Awal," *Pengabd. Kpd. Masy.*, vol. 1, pp. 6–12, 2017.
- [4] N. . Pieter, H.Z Lubis, *Pengantar Psikologi Dalam Perawatan*. Jakarta, 2010.
- [5] W. S. Perbowosari, H Indrawan, I Hadion, *Pengantar Psikologi Pendidikan*. Jawa Timur: Qiara Media, 2019.
- [6] W. A. Laksono, Y. Astuti, S. Informasi, U. A. Yogyakarta, M. Informatika, and U. A. Yogyakarta, "Metode Myer Briggs Type Indicator (MBTI) untuk Tes Kepribadian Sebagai Media Pengembangan Diri (Studi Kasus : SMAN 2 Kebumen) Pendahuluan Hasil dan Pembahasan Tinjauan Pustaka Metode Penelitian," vol. 1, no. 2, pp. 1–6, 2020.
- [7] F. Littauer, *Personality Plus*. 1992.
- [8] Bustami, "Penerapan Algoritma Naive Bayes untuk Mengklasifikasi Data Nasabah Asuransi," *Informatika*, vol. 6, 2014.
- [9] I. Muhammad, H Prasojo, C.A Sugianto, N.A Surtiningsih, L Cholissodin, "Optimasi Naive Bayes Classifier dengan Menggunakan Particle Swarm Optimization pada Data Iris," *Tekno. Inf. dan Ilmu Kompter*, vol. 4, pp. 180–184, 2017.
- [10] R. Wati, "Penerapan Algoritma Genetika untuk Seleksi Fitur pada Analisis Sentimen Review Jasa Maskapai Peerbangan Menggunakan Naive Bayes," *Evolusi*, vol. 4, 2016.
- [11] H. . Saputra, "Analisis Data Mining untuk Pemetaan Mahasiswa yang Membutuhkan Bimbingan dan Konseling Menggunakan Algoritma Naive Bayes Classifier," *Tekno. Inf. dan Pendidik.*, vol. 11, 2018.
- [12] A. Nasution, N Djahara, K Zamsure, "Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes (Studi Kasus: Fasilkom Unilak)," *Tekno. Inf. dan Komun. Digit. Zo.*, vol. 6, pp. 1–11, 2015.
- [13] S. Ja'far, "Struktur Kepribadian Manusia Perspektif Psikologi dan Filsafat," *Ilm. Psikol.*, vol. 2, pp. 209–221, 2015.
- [14] G. . Pranata, H.K Subari Gunawan, "Penerapan Metode Naive Bayes untuk Klasifikasi SMS Spam Menggunakan Java Programming," *J-Intech*, vol. 7, 2019.
- [15] M. . Putri, *Pengaruh Kepribadian Menurut Teori Big Five Factor Terhadap Sikap Altruisme pada Mahasiswa Bina Nusantara*. 2015.
- [16] Kuntjojo, *Psikologi Kepribadian*. Kediri, 2009.
- [17] Y. Mardi, "Data Mining: Klasifikasi Menggunakan Algoritma C4.5," *Edik Inform.*, vol. 2, 2019.