

## Penerapan Algoritma *Naive Bayes* untuk Klasifikasi Tingkat Kebugaran Jasmani Berdasarkan Hasil Pengukuran pada Pegawai

Alleyda Irzky Shafarindu<sup>1</sup>, lin Ernawati, S.Kom., M.Si.<sup>2</sup>, Ati Zaidiah, S.Kom., MTI.<sup>3</sup>  
Program Studi Informatika, Universitas Pembangunan Nasional Veteran Jakarta  
Jl. RS. Fatmawati Raya, Pd. Labu, Kec. Cilandak, Kota Depok, Jawa Barat 12450  
alleydais@upnvj.ac.id<sup>1</sup>, iinernawati@upnvj.ac.id<sup>2</sup>, atizaidiah@upnvj.ac.id<sup>3</sup>

**Abstrak.** Dalam melakukan suatu aktivitas, dibutuhkan kebugaran jasmani agar aktivitas dapat dilakukan secara optimal dan efisien. Kebugaran jasmani memiliki peranan penting yang dapat mempengaruhi produktivitas kerja. Dalam rangka hari kesehatan nasional, Dinas Kesehatan Provinsi Kepulauan Bangka Belitung melaksanakan pengukuran kebugaran jasmani untuk mengetahui tingkat kebugaran jasmani para pegawai, pengukuran dilakukan dengan Tes *Rockport* atau Tes Jalan 6 menit bagi pegawai yang memiliki penyakit yang berisiko tinggi. Pada penelitian ini, akan dilakukan klasifikasi tingkat kebugaran jasmani berdasarkan hasil pengukuran kebugaran jasmani pada pegawai. Klasifikasi dilakukan menggunakan algoritma *Naive Bayes*. Algoritma *Naive Bayes* merupakan salah satu algoritma yang dapat digunakan dalam metode klasifikasi. Pada penelitian ini dilakukan pembagian data dengan menerapkan *K-fold cross validation* dengan nilai  $k = 4$ . Dari penelitian ini dihasilkan nilai akurasi sebesar 94%, nilai presisi sebesar 92%, nilai *recall* sebesar 94%, dan nilai *F1-Score* sebesar 93%.

**Kata kunci:** pengukuran Kebugaran jasmani, klasifikasi, *naive bayes*.

### 1 Pendahuluan

Kebugaran merupakan salah satu faktor penting dalam kehidupan sehari-hari. Saat melakukan suatu aktivitas, Anda harus berada dalam kondisi fisik yang baik agar dapat melakukan aktivitas tersebut secara optimal dan efektif. Kebugaran fisik diperlukan untuk berfungsi dengan baik tanpa kelelahan yang serius. Untuk mengetahui tingkat kebugaran jasmani seseorang, diperlukan pengukuran kebugaran jasmani. Pengukuran kebugaran jasmani merupakan suatu upaya untuk mengukur tingkat kebugaran seseorang yang dapat menjadi dasar untuk menentukan program latihan fisik yang dibutuhkan untuk meningkatkan kebugaran orang tersebut. Tingkat kebugaran yang rendah dapat mempengaruhi produktivitas dan kinerja. [1]

Dalam rangka hari kesehatan nasional, Dinas Kesehatan Provinsi Kepulauan Bangka Belitung melaksanakan pengukuran kebugaran jasmani untuk mengetahui tingkat kebugaran jasmani para pegawai. Pengukuran ini ditujukan untuk membuat kebugaran ASN meningkat sehingga mewujudkan ASN yang produktif serta menerapkan gaya hidup sehat. Untuk mempersiapkan ASN yang bugar dan produktif, maka Kementerian Kesehatan RI melalui program kesehatan kerja dan olahraga mendukung untuk pengukuran kebugaran jasmani bagi pegawai menggunakan Tes *Rockport* dengan jarak 1,600 meter atau Tes Jalan 6 menit bagi pegawai yang memiliki penyakit yang berisiko tinggi.

Data mining merupakan salah satu proses untuk melakukan pengolahan data yang bertujuan untuk mengekstrak atau mengambil inti sari dari pengetahuan dan sekumpulan data sehingga didapatkan struktur yang dapat lebih mudah dimengerti. Dalam penelitian ini, metode data mining yang digunakan adalah klasifikasi. Klasifikasi merupakan pengelompokan berdasarkan hubungan antara variable kriteria dengan variabel target. Dalam proses klasifikasi terdapat beberapa algoritma yang dapat digunakan. Pada penelitian ini, algoritma yang akan digunakan yaitu algoritma *Naive Bayes*. Algoritma ini dipilih karena mempunyai nilai akurasi yang cukup tinggi dibandingkan dengan algoritma lainnya. Pada penelitian ini akan dilakukan penerapan salah satu metode data mining yaitu metode klasifikasi untuk mendapatkan hasil klasifikasi untuk mengetahui tingkat kebugaran jasmani dari hasil pengukuran dengan menggunakan algoritma *Naive Bayes* dengan hasil berupa hasil klasifikasi.

## 2 Tinjauan Pustaka

### 2.1 Kebugaran Jasmani

Kebugaran jasmani merupakan kemampuan fisik seseorang untuk beradaptasi terhadap tekanan fisik tanpa menimbulkan kelelahan yang berarti dalam pekerjaan sehari-hari. Seseorang dengan kebugaran jasmani yang baik, dapat meningkatkan kinerjanya dan mengurangi risiko cedera dalam aktivitas sehari-hari. Kondisi fisik seseorang berkaitan dengan kemampuan fisik seseorang untuk melakukan pekerjaan secara optimal dan efisien. [2]

### 2.2 Pengukuran Kebugaran Jasmani

Pengukuran kebugaran jasmani merupakan suatu usaha untuk mengukur tingkat kebugaran seseorang yang bisa menjadi dasar dalam penentuan program latihan fisik untuk peningkatan kebugaran jasmani orang tersebut. Dengan adanya pengukuran kebugaran jasmani diharapkan pegawai dapat mengetahui tingkat kebugarannya, sehingga termotivasi untuk menjadi lebih sehat, bugar, dan produktif.[1]

Ada banyak metode tes yang dapat dilakukan untuk melakukan pengukuran kebugaran jasmani. Salah satunya dengan metode melakukan *single test rockport* 1600 meter untuk dewasa dengan umur diatas 19 tahun serta tes jalan 6 menit untuk dewasa dengan umur diatas 60 tahun atau dengan kondisi risti atau kondisi dimana seseorang potensi untuk sakit. Berikut merupakan metode pengukuran kebugaran jasmani yang dilakukan pada data yang didapatkan untuk penelitian ini.

#### 2.2.1 Metode Tes *Rockport*

Metode tes *rockport* merupakan salah satu metode untuk mengukur tingkat kebugaran jasmani. Pada tes *rockport*, peserta diharuskan untuk melakukan jalan atau lari sesuai dengan kemampuan masing-masing dengan konstan pada lintasan yang datar atau landai sejauh 1.600 meter. Waktu tempuh yang dicapai peserta dalam menit dan detik yang didapat akan dikonversi dengan tabel VO2Max dan tabel usia yang akan menunjukkan tingkat kebugaran peserta berdasarkan kemampuan jantung paru. [1]

#### 2.2.2 Metode jalan 6 menit

Tes jalan 6 menit merupakan salah satu metode pengukuran kebugaran jasmani yang menjadi alternatif metode pengukuran bagi peserta yang memiliki penyakit yang berisiko tinggi. Tes jalan 6 menit diperuntukkan bagi peserta yang tidak lulus skrining 7 pertanyaan atau *Par-Q and You Test*. Pertanyaan pada *Par-Q and You Test*. Pada proses pengukuran, peserta diharuskan berjalan atau jalan cepat selama 6 menit. Hasil jarak dalam satuan meter akan dikonversikan sesuai usia dan jenis kelamin kedalam tabel. Sehingga akan didapat kategori seperti tes *rockport*. [1]

### 2.3 Data Mining

Data mining merupakan kombinasi beberapa cabang ilmu komputer yang merupakan sebuah proses menemukan pola pada suatu kumpulan data yang besar. Metode penemuan termasuk metode irisan dari *artificial intelligence*, *machine learning*, *statistics*, dan *database system*. Tujuan dari data mining yaitu mengambil intisari dari suatu pengetahuan untuk mengumpulkan data dan menghasilkan struktur data yang lebih mudah dimengerti. Data mining merupakan metode analisis dari proses pencarian pengetahuan pada basis data atau biasa disebut *knowledge discovery in databases* (KDD). Pengetahuan tentang masalah dapat berbentuk model data, atau hubungan antara data yang valid atau yang sebelumnya tidak diketahui. Data mining adalah proses menemukan pengetahuan, pola, dan hubungan antar suatu data secara otomatis. Pengetahuan dapat ditemukan dengan lima proses berurutan yaitu seleksi, pra pemrosesan, transformasi, data mining, dan interpretasi atau evaluasi.[3]

## 2.4 Klasifikasi

Klasifikasi merupakan metode yang digunakan untuk menetapkan *record* data baru ke salah satu kelas atau kategori yang telah ditentukan sebelumnya. Dengan kata lain klasifikasi juga berguna untuk memprediksi label ditiap kategori kelas. Fungsi lainnya adalah mengklasifikasikan (memodelkan) data berdasarkan set data *training* dan nilai (label kelas) dalam atribut klasifikasi untuk digunakan untuk mengklasifikasikan data yang baru.[4]

Proses klasifikasi data memiliki dua tahapan, yaitu *learning* yang berarti pada tahap data *training* dianalisa menggunakan algoritma klasifikasi. Lalu *classification* yang berarti pada tahap ini data test akan digunakan untuk mengestimasi ketepatan dari *classification rules*. Jika presisi yang disesuaikan dan yang diprediksikan diterima, *rule* yang didapatkan dapat diterapkan pada klasifikasi lainnya dari *tuple* data yang baru. [5]

## 2.5 Algoritma Naïve Bayes

Teorema Bayesian adalah pendekatan statistik dasar untuk pengenalan pola. Teorema Bayes digunakan untuk melakukan perhitungan probabilitas dari suatu hipotesis dari probabilitas *prior* yang diberikan. Probabilitas *prior* merupakan nilai probabilitas yang dipercayai benar dari awal, sebelum dilakukan pengamatan dan analisis terhadap suatu data. Oleh karena itu, algoritma dasar untuk melakukan perhitungan probabilitas pada setiap kemungkinan hipotesis untuk penentuan kemungkinan terdekat dengan menggunakan Teorema Bayes. Teorema bayes dinamai Thomas Bayes, yang menemukan teorema bayes pada abad ke-18. Metode bayesian menggunakan probabilitas bersyarat sebagai dasarnya. Berikut adalah persamaan teorema Bayes. [3]

$$P(X) = \frac{P(H)P(H)}{P(X)} \quad (1)$$

Keterangan:

P = Probabilitas

H = Hipotesis

X = Bukti

$P(H|X)$  = Probabilitas bahwa hipotesis H benar untuk bukti X

$P(X|H)$  = Probabilitas bahwa bukti X benar untuk hipotesis H

$P(H)$  = Probabilitas *prior* hipotesis H

$P(X)$  = Probabilitas *prior* bukti X

## 2.6 Evaluasi Kinerja Klasifikasi

Evaluasi *classifier* biasanya dilakukan dengan menggunakan sebuah kumpulan data uji dengan ukuran tertentu yang tidak digunakan untuk melatih *classifier*. Ada berbagai ukuran yang bisa digunakan untuk mengevaluasi model klasifikasi, di antaranya adalah *accuracy* atau tingkat pengenalan, *recall* atau *sensitivity* atau *true positive rate*, dan *F1-score* atau rata-rata harmonik dari *precision* dan *recall*. Berikut ukuran evaluasi kinerja model klasifikasi. [3]

**Tabel 1.** Ukuran Evaluasi Kinerja Klasifikasi

No.	Ukuran	Rumus
1	Accuracy atau tingkat pengenalan	$\frac{TP + TN}{P + N}$
2	Recall atau sensitiviti atau true positive rate	$\frac{TP}{P}$
3	Presisi atau Precision	$\frac{TP}{TP + FP}$
4	F1-Score atau rata-rata harmonik dari presicion dan recall	$2 \times \frac{precision \times recall}{precision + recall}$

## 2.7 Studi Literatur

Studi literatur yang digunakan pada penelitian ini merupakan hasil penelitian yang sudah dibuat sebelumnya sebagai bahan referensi. Studi literatur yang digunakan sebagai berikut.

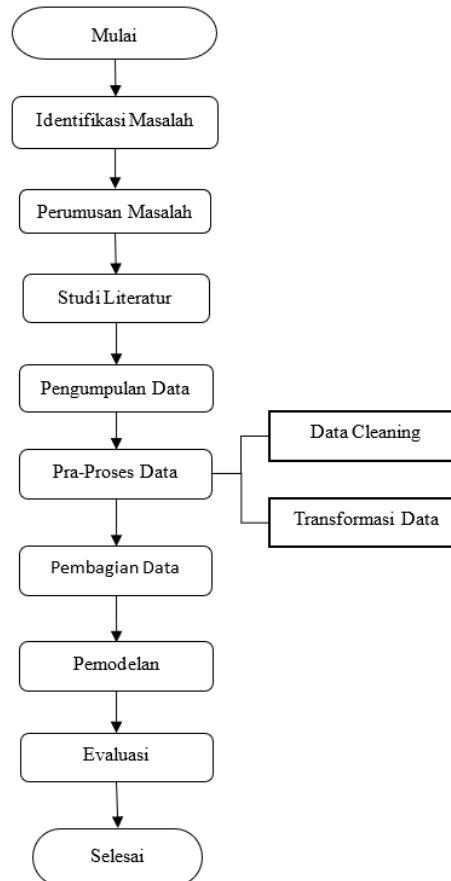
Anis Zubair dan Moch Muksin (2018) melakukan penelitian dengan judul “Penerapan Metode *Naive Bayes* Untuk Klasifikasi Status Gizi (Studi Kasus Di Klinik Bromo Malang)”. Penelitian ini dirancang untuk mengklasifikasikan status gizi. Penelitian ini digunakan data mining untuk klasifikasi dengan menerapkan algoritma *Naive Bayes*. Data yang digunakan sebanyak 49 sampel yang didapat dari klinik gizi yang ada di kota Malang yaitu Klinik Bromo Pada pengolahan data dilakukan penghapusan beberapa atribut yaitu umur, berat badan, lingkar perut, dan *visceral fat*. Berdasarkan hasil penelitian ini, klasifikasi status gizi menggunakan metode *Naive Bayes* menghasilkan nilai akurasi yang baik dengan nilai sebesar 98%.

Muhammad Uli Nuha (2020) melakukan penelitian dengan judul “Klasifikasi Status Gizi Balita Menggunakan Metode *Naive Bayes* Classifier Berbasis Website”. Penelitian ini dirancang untuk memudahkan penentuan status gizi pada bayi dibawah lima tahun atau balita dengan data mining. Penelitian ini menerapkan metode klasifikasi dengan menggunakan algoritma *Naive Bayes*. Penelitian menggunakan 21 contoh dataset status gizi balita. Berdasarkan hasil penelitian ini, sistem klasifikasi status gizi yang telah dibuat mendapatkan nilai akurasi yang baik dengan nilai sebesar 90% dan galat sebesar 10% dengan menggunakan pengujian *confusion matrix*.

Nani Purwati (2018) melakukan penelitian dengan judul “Deteksi Gizi Buruk Pada Balita Berdasarkan Indeks Antropometri Menggunakan Algoritma *Naive Bayes*”. Penelitian ini dirancang untuk melakukan uji coba untuk deteksi gizi buruk pada balita berdasarkan indeks antropometri baik berdasarkan indeks berat badan (BB)/ Umur (U) dan BB/TB (Tinggi Badan). Penelitian ini menerapkan algoritma *Naive Bayes*. Data yang digunakan berjumlah 261 data balita dibawah umur 5 tahun (0-59 bulan). Berdasarkan hasil penelitian ini, dihasilkan akurasi yang baik dengan nilai sebesar 90,20%.

Rahmatina Hidayati, Nuryuliana, dan Anis Zubair (2020) melakukan penelitian dengan judul “Klasifikasi Status Kesehatan Penerimaan Calon Karyawan Menggunakan Metode *Naive Bayes*”. Penelitian ini dirancang untuk melakukan klasifikasi untuk mengetahui status kesehatan calon karyawan pada suatu perusahaan yang dalam memilih karyawan diperlukan penentuan kualitas calon karyawan melalui status kesehatan. Penelitian ini menerapkan algoritma *Naive Bayes* dalam melakukan klasifikasi. Data yang digunakan sebanyak 56 data hasil pemeriksaan medis dengan aspek-aspek yang diperiksa yaitu tinggi badan, berat badan, tekanan darah, mata, dan sejarah penyakit kronis. Dari penelitian ini, klasifikasi status kesehatan penerimaan calon karyawan potensial perusahaan “X” menggunakan metode *Naive Bayes* menghasilkan nilai akurasi sebesar 89%.

### 3 Metode Penelitian



**Gambar 1.** Alur Penelitian

#### 3.1 Identifikasi Masalah

Pada langkah ini dilakukan identifikasi masalah. Dari langkah ini dilakukan pengamatan. Pengamatan yang didapat ialah seseorang yang mempunyai kebugaran jasmani yang baik bisa melakukan suatu aktivitas lebih optimal sehingga dapat meningkatkan produktivitas manusia sehingga aktivitas yang dilakukannya dapat berjalan dengan baik sehingga dapat melaksanakan tugas dan fungsinya.

#### 3.2 Perumusan Masalah

Tahap perumusan masalah adalah tahap dimana pada penelitian telah disusun masalah yang telah ditemukan pada tahap identifikasi masalah. Rumusan masalah ialah mengenai bagaimana penerapan data mining dalam melakukan klasifikasi menggunakan algoritma *Naive Bayes* terhadap hasil pengukuran kebugaran jasmani pada pegawai.

### 3.3 Studi Literatur

Studi literatur yang dilakukan adalah dengan melakukan pengumpulan data dan informasi yang dapat mendukung proses penulisan penelitian ini yang berkaitan dengan judul penelitian. Sumber studi literatur didapat dari berbagai sumber. Diantaranya dari buku, web dan jurnal-jurnal yang terkait dengan kebugaran jasmani, pengukuran jasmani, data mining, klasifikasi, *Naive Bayes* serta alat bantu yang digunakan dalam proses pengolahan data yang telah didapatkan.

### 3.4 Pengumpulan Data

Pada tahap ini dilakukan pengumpulan data untuk penelitian. Data yang digunakan dalam penelitian ini adalah dataset yang didapat dari hasil pengukuran kebugaran jasmani pegawai di Dinas Kesehatan Provinsi Kepulauan Bangka Belitung pada tahun 2019 dan 2020. Dataset yang diberikan berjumlah 201 data yang terdiri dari 10 variabel yang terdiri dari 9 fitur dan 1 kelas. Variabel fitur yang digunakan yaitu Jenis Kelamin, Umur, Tinggi Badan, Berat Badan, IMT, Status Gizi, Metode, Waktu Tempuh/Jarak Tempuh, dan VO2 Max. Kelas pada data ini berada pada Hasil Pengukuran yaitu “Kurang Sekali”, “Kurang”, “Cukup”, “Baik”, dan “Baik Sekali”.

### 3.5 Pra-Proses Data

Tahap pra-proses data untuk menjadikan data menjadi suatu data yang dapat digunakan dan mudah untuk dibaca (*readable*). Tahap-tahap yang ada pada pra-proses penelitian ini sebagai berikut.

#### 3.5.1 Data Cleaning

*Data cleaning* atau pembersihan data dilakukan dengan melihat data yang digunakan apakah terdapat informasi (variabel) yang hilang atau tidak yang biasa disebut dengan *missing value*. Serta melakukan reduksi data pada variabel yang tidak digunakan untuk memastikan data pada hasil pengukuran yang dipilih dapat digunakan untuk proses pengolahan dan diproses menggunakan algoritma *Naive Bayes*.

#### 3.5.2 Transformasi Data

Pada tahap ini dilakukan transformasi data yang dilakukan agar data yang digunakan sesuai dengan algoritma yang akan digunakan. Variabel yang berjenis kategorikal yaitu Jenis Kelamin, Status Gizi, dan Metode akan dilakukan proses inisialisasi ke dalam bentuk numerical

### 3.6 Pembagian Data

Tahapan ini merupakan tahapan dilakukan pembagian data. Pembagian data dilakukan dengan metode *k-fold cross validation*. Pada pembagian data akan dilakukan 4 kali percobaan pada nilai K 2 sampai 5 untuk mencari hasil klasifikasi yang lebih baik. *K-fold cross validation* merupakan salah satu metode untuk memisahkan data training dan data testing. *K-fold cross validation* membagi suatu kumpulan data secara acak menjadi *k fold* atau sub-kumpulan yang saling bebas sehingga setiap pemisahan fold berisi  $1/k$  data. Dalam metode *k-fold cross validation*, setiap sampel dalam data latih dan setiap sampel memiliki jumlah kemunculan hanya satu kali dalam data uji.[3]

### 3.7 Pemodelan

Tahapan ini merupakan tahapan dilakukan pemodelan untuk melakukan klasifikasi dari dataset yang telah di pra-proses sebelumnya. Pemodelan dilakukan dengan menerapkan algoritma *Naive Bayes*. Proses ini dilakukan dengan membuat *code* dengan menggunakan Bahasa python. Pada pemodelan akan dilakukan *input* dataset yang akan digunakan. Setelahnya akan dilakukan pra-proses pada data tersebut. Lalu akan dilakukan pembagian data menjadi

data uji dan data latih dengan menggunakan *k-cross validation*. Setelah itu akan dilakukan pembangunan model dengan memanggil library *Naive Bayes Gaussian*. Kemudian akan dilakukan pembuatan *Gaussian Classifier*. Lalu dilakukan pelatihan model dengan data training. Setelahnya akan dilakukan penentuan hasil prediksi dari data testing dan probabilitas hasil prediksi. Lalu mencari hasil *confusion matrix* dan evaluasi.

### 3.8 Evaluasi

Pada tahap ini merupakan tahapan evaluasi dari hasil pemodelan yang telah diterapkan. Untuk mengetahui tingkat hasil klasifikasi pada pemodelan yang telah dibuat, akan dilakukan pengukuran performa dengan melihat nilai akurasi, *recall*, presisi, dan *F1-Score*.

## 4 Hasil dan Pembahasan

### 4.1 Data

Tahap pertama dalam penelitian ini adalah pengumpulan data hasil kebugaran jasmani. Data hasil kebugaran jasmani ini didapatkan dengan melakukan tes kebugaran jasmani pada pegawai yang dilakukan oleh Dinas Kesehatan Provinsi Kepulauan Bangka Belitung. Tes ini dilaksanakan dengan dua metode yaitu metode tes *rockport* atau tes jalan 6 menit. Pada tes *rockport* akan dihitung waktu tempuh untuk mencapai 1600 meter. Sedangkan pada tes jalan 6 menit akan dihitung jarak tempuh yang berhasil dilalui dalam waktu 6 menit. Kelas pada data ini berada pada variabel Hasil Pengukuran yang terdiri dari yaitu “Kurang Sekali”, “Kurang”, “Cukup”, “Baik”, dan “Baik Sekali”. Dari hasil pengukuran 64 orang mendapatkan hasil “Kurang Sekali”, 42 orang mendapatkan hasil “Kurang”, 74 orang mendapatkan hasil “Cukup”, 18 orang mendapatkan hasil “Baik”, dan 3 orang mendapatkan hasil “Baik Sekali”. Berikut merupakan keterangan dari variabel yang ada pada data.

**Tabel 2.** Data

No.	Variabel	Keterangan	Jenis	Nilai
Fitur				
1	Jenis Kelamin	Jenis Kelamin dari peserta tes	Kategorikal	1. Laki-laki 2. Perempuan
2	Umur	Umur dari peserta tes dalam tahun	Numerik	26 - 58
3	Tinggi Badan	Tinggi badan peserta tes dalam cm	Numerik	144 - 179
4	Berat Badan	Berat badan peserta tes dalam kg	Numerik	42 - 89.5
5	IMT	Indeks Massa Tubuh dalam kg/m <sup>2</sup>	Numerik	17,85 – 36,20
6	Status Gizi	Status gizi	Kategorikal	1. Kurus 2. Normal 3. Gemuk 4. Obesitas
7	Metode	Metode tes yang dilaksanakan	Kategorikal	1. Rockport 2. Jalan 6 menit
8	Waktu Tempuh / Jarak Tempuh	Waktu peserta metode tes rockport / Jarak tempuh peserta dengan tes metode jalan 6 menit. Waktu dalam menit / Jarak dalam meter.	Numerik	Waktu: 6,50 – 21,00  Jarak: 60 -800
9	VO2 MAX	Perhitungan VO2 Max.	Numerik	20 - 50

		(Terdapat Missing Value)		
Kelas				
1	Hasil Pengukuran	Hasil pengukuran kebugaran jasmani	Kategorikal	1. Kurang Sekali 2. Kurang 3. Cukup 4. Baik 5. Baik Sekali

## 4.2 Pra-proses Data

### 4.2.1 Cleaning Data

*Data cleaning* atau pembersihan data dilakukan dengan melihat data yang digunakan apakah terdapat informasi (variabel) yang hilang atau tidak yang biasa disebut dengan *missing value*. Pada atribut VO2 Max terdapat *missing value* (NaN), *missing value* ada pada data dari peserta tes jalan 6 menit. Karena pada tes jalan 6 menit yang diukur adalah jarak tempuh saja tanpa menghitung VO2 Max. Dari total 201 data, terdapat 79 data yang memiliki nilai NaN. Karena jumlah data yang memiliki missing value cukup besar, maka variabel VO2 Max dilakukan penghapusan. Sebelumnya dilakukan tiga percobaan dengan memproses data yang missing value dari variabel VO2 Max diisi menjadi 0, data dengan *missing value* diisi dengan nilai rata-rata dari variabel VO2 Max, dan dengan data yang variabel VO2 Max dihapuskan. Ketiganya mendapatkan hasil klasifikasi yang sama. Sehingga pada penelitian ini dilakukan reduksi atau penghapusan variabel VO2 Max.

### 4.2.2 Transformasi Data

Pada tahap ini dilakukan transformasi data agar data yang digunakan dapat dilakukan pemrosesan. Pada tahap ini variabel fitur yang berjenis kategorikal akan diubah menjadi numerikal. Dikarenakan adanya tidak keseimbangan data, maka pada variabel kelas hasil pengukuran akan diubah dari lima kelas menjadi dua kelas. Untuk kelas “Kurang Sekali” dan “Kurang” akan diubah menjadi “Tidak Bugar”. Karena dari hasil yang didapatkan para peserta dengan kelas tersebut dianggap Tidak Bugar. Untuk kelas “Cukup”, “Baik”, dan “Baik Sekali” akan diubah menjadi kelas “Bugar”. Karena dari hasil yang didapatkan para peserta dengan kelas tersebut dianggap Bugar.

## 4.3 Pembagian Data

Pembagian data dilakukan untuk membagi data menjadi data *testing* dan data *training*. Data latih akan digunakan oleh algoritma klasifikasi untuk menciptakan sebuah model klasifikasi, sedangkan data uji akan digunakan untuk menguji performa dan kebenaran model dalam melakukan klasifikasi. Dalam penelitian ini, pembagian data latih dan data uji dilakukan dengan menggunakan metode *K-Fold Cross Validation*. Sebelum dilakukan pembagian data, fitur dan kelas akan dipisahkan, X untuk fitur dan Y untuk kelas. Setelah dilakukan beberapa percobaan, nilai k yang diambil adalah 4 untuk data dengan kelas “Bugar” dan “Tidak Bugar”. Sehingga akan digunakan 151 data *training* dan 50 data *testing*. Dibawah ini merupakan jumlah pembagian data pada data testing untuk setiap kelas dengan menggunakan nilai k adalah 3 dan 4 untuk data dengan 2 kelas. Berikut merupakan pembagian data *testing* untuk 2 kelas yaitu “Bugar” dan “Tidak Bugar” menggunakan nilai k adalah 3 dan 4.

**Tabel 3.** Pembagian data *testing* pada 2 kelas

Nilai k	Kelas		Jumlah
	Bugar	Tidak Bugar	
k = 3	16	51	67
k = 4	15	35	50



#### 4.4 Pemodelan

Pemodelan dilakukan dengan menggunakan library dari scikit learn yaitu *Gaussian Naive Bayes*. Pada pemodelan akan dilakukan pemanggilan fungsi klasifikasi *naive bayes*. Lalu data *training* akan dimasukkan pada fungsi *GaussianNB*. Setelahnya akan dilakukan hasil prediksi dari data testing untuk menentukan probabilitas hasil prediksi. Fungsi *GaussianNB* mengimplementasi algoritma untuk klasifikasi. Pada penelitian ini digunakan *Gaussian naive bayes* karena data yang digunakan merupakan data kontinu. Serta seperti studi literature yang digunakan, menggunakan *Gaussian naive bayes* dan mendapatkan hasil akurasi yang baik. Sehingga pada penelitian ini akan digunakan *Gaussian naive bayes*.

#### 4.5 Evaluasi Hasil

Dilakukan berbagai percobaan dalam melakukan penelitian ini. Dengan menggunakan data dengan lima kelas dan dua kelas serta menggunakan dua metode pembagian data yaitu *K-Fold Cross Validation* dan *Train Test Split*. Berikut merupakan hasil akurasi dari beberapa percobaan yang telah dilakukan. Hasil percobaan menggunakan data dengan lima kelas yaitu “Kurang Sekali”, “Kurang”, “Cukup”, “Baik”, dan “Baik Sekali”. Nilai akurasi terbesar ada pada pembagian data *K-Fold* dengan nilai k adalah 3 yaitu 0.75 atau 75%. Hasil percobaan menggunakan data dengan dua kelas yaitu “Bugar” dan “Tidak Bugar”. Nilai akurasi terbesar ada pada pembagian data *K-Fold* dengan nilai k adalah 3 dan 4 yaitu 0.94 atau 94%. Sehingga pembagian data yang digunakan merupakan metode *K-Fold Cross Validation* dengan nilai k adalah 4. Hasil klasifikasi yang dihasilkan cukup memuaskan. Tabel hasil klasifikasi dari data kebugaran jasmani sebagai berikut.

**Tabel 4.** Hasil klasifikasi data kebugaran jasmani

	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>	<i>Support</i>
Bugar	0.88	0.93	0.90	15
Tidak Bugar	0.97	0.94	0.96	35
<i>Accuracy</i>	0.94			50
<i>Average</i>	0.92	0.94	0.93	50

Dari tabel hasil klasifikasi, Hasil klasifikasi yang didapatkan memiliki nilai-nilai yang baik. Dapat dilihat jika nilai akurasi yang dihasilkan dari klasifikasi yang dilakukan sebesar 0.94 atau 94%. Hal ini menyatakan bahwa terdapat 94% jumlah tuple dalam data uji yang diklasifikasikan dengan benar oleh *classifier*. Nilai *precision* pada klasifikasi mendapatkan hasil sebesar 0.92 atau 92%, yaitu terdapat 92% tuple yang dilabeli sebagai positif adalah benar pada hasil prediksi yang dilakukan. Lalu klasifikasi yang dilakukan mendapatkan nilai *recall* sebesar 0.94 atau 94% yaitu terdapat 94% tuple positif yang diklasifikasikan dengan benar pada keseluruhannya. Nilai *F1 Score* yang dihasilkan sebesar 0.93 atau 93%, 0.90 untuk kelas bugar dan 0.96 untuk kelas tidak bugar.

Adapun ciri data untuk kelas Bugar adalah untuk metode test *rockport* memiliki nilai waktu yang kecil dan nilai VO2 Max yang besar sedangkan untuk metode test jalan 6 menit memiliki nilai jarak yang besar. Kemudian ciri data untuk kelas tidak bugar adalah untuk metode test *rockport* memiliki nilai waktu yang besar dan nilai VO2 Max yang kecil sedangkan untuk metode test jalan 6 menit memiliki nilai jarak yang kecil.

## 5 Kesimpulan dan Saran

Setelah melakukan penelitian pada data kebugaran jasmani dengan menggunakan algoritma *naive bayes* maka dapat disimpulkan bahwa penerapan data mining menggunakan algoritma *Naive Bayes* dalam melakukan klasifikasi tingkat kebugaran jasmani pada pengolahan data dilakukan pra-proses dengan menghapus variabel VO2 Max karena memiliki *missing value* yang cukup banyak. Lalu dilakukan transformasi data kategorikal menjadi numerikal dan mengubah kelas dari 5 kelas yaitu “Kurang Sekali”, “Kurang”, “Cukup”, “Baik”, dan “Baik Sekali” menjadi “Bugar” dan “Tidak Bugar”. Hasil klasifikasi yang diperoleh dari pengolahan data pengukuran kebugaran jasmani pegawai dengan menerapkan algoritma *Naive Bayes* yaitu memiliki nilai akurasi sebesar 94%, nilai presisi sebesar 92%, nilai *recall* sebesar 94%, dan nilai *F1-Score* sebesar 93%.

Saran untuk penelitian ini diantaranya meningkatkan jumlah data yang lebih banyak dengan mencari data di perkantoran lain atau menambah jumlah peserta dan variabel yang lebih spesifik seperti denyut nadi dari peserta dan gunakan atau bandingkan dengan algoritma lainnya seperti *Decision Tree* dan *K-Nearest Neighbor*.

## Referensi

- [1] Kementerian Kesehatan RI. (2017). *Petunjuk Teknik Pengukuran Kebugaran Jasmani Terintegrasi Bagi Aparatur Sipil Negara Kementerian Kesehatan*.
- [2] Adi, S., Supriyadi, & Masgumelar, N. K. (2020). *Model-Model Exercise Dan Aktivitas Fisik Untuk Kebugaran Jasmani Anak SD*. Wineka Media.
- [3] Suyanto. (2017). *Data Mining Untuk Klasifikasi dan Klasterisasi Data*. Penerbit Informatika.
- [4] Prasetyowati, E. (2017). *DATA MINING Pengelompokan Data untuk Informasi dan Evaluasi*. Duta Media Publishing.
- [5] Vulandari, R. T. (2017). *Data Mining Teori dan Aplikasi RapidMiner*. Penerbit Gava Media.