

Terakreditasi SINTA Peringkat 4

Surat Keputusan Dirjen Penguatan Riset dan Pengembangan Ristek Dikti No. 28/E/KPT/2019
masa berlaku mulai Vol.3 No. 1 tahun 2018 s.d Vol. 7 No. 1 tahun 2022

Terbit online pada laman web jurnal:
<http://publishing-widyagama.ac.id/ejournal-v2/index.php/jointecs>



Vol. 5 No. 3 (2020) 177 - 184

JOINTECS

(Journal of Information Technology and Computer Science)

e-ISSN:2541-6448

p-ISSN:2541-3619

Prediksi Status Konsumen Produk Celana Menggunakan *Naïve Bayes*

Din Syamsudin¹, Yosia Christmas Decky Halundaka², Aryo Nugroho³

Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Narotama

¹din.syamsudin@fik.narotama.ac.id, ²yosia.christmas@mhs.fasilkom.narotama.ac.id,

³aryo.nugroho@narotama.ac.id

Abstract

Pants material is a daily staple, a problem faced by buyer among students, employees, and college students. This aims of this study to predict the Consumer using the Naïve Bayes method. Because the Naïve Bayes method is considered good for classification and accuracy compared to the C4.5 Algorithm method has instability in doing classification. Retrieval of data in this study from the results of interviews with sellers in January to December 2019 then conducted input in Excel with a total of 731 data. By selecting data to get perfect and accurate results, trimming the variables with the 10 initial variables into 5 variables. In the preprocessing stage before conducting the data mining process to find out the amount of data per attribute then the data mining process uses the naïve Bayes algorithm, with a combination of 60 to 90 percent split percentage showing the results of each accuracy value 80.137%, 78.0822%, 81, 5060% and 83.5616%. The best is the 90% split percentage with an accuracy percentage of 83.5616%. The results of the evaluation using the confusion matrix shows that the highest and superior consumer status of the employee class and student class is to show the college student class.

Keywords: *prediction; classification; naïve bayes; data mining; sales*

Abstrak

Celana bahan merupakan kebutuhan pokok sehari-hari, permasalahan yang dihadapi ketidak pastian pembeli dikalangan antara Mahasiswa, Karyawan, dan Pelajar. Oleh karena itu penelitian ini bertujuan untuk memprediksi Status Konsumen Produk Celana menggunakan metode *Naïve Bayes*. karena metode *Naïve Bayes* yang dinilai baik melakukan klasifikasi dan akurasi dibandingkan metode *Algoritma C4.5* yang memiliki ketidakstabilan dalam melakukan klasifikasi. Pengambilan data pada penelitian ini dari hasil wawancara pada penjual dibulan januari hingga desember tahun 2019 kemudian dilakukan penginputan pada excel dengan total 731 data. Dengan menseleksi data agar mendapatkan hasil yang sempurna dan akurat dilakukan pemangkasan variable dengan awal 10 variable menjadi 5 variable. Pada tahap *preprocessing* sebelum melakukan proses data mining untuk mengetahui jumlah data per-*attribute* lalu melakukan proses data mining menggunakan algoritma *naïve bayes*, dengan *percentage split* kombinasi 60 hingga 90 persen menunjukkan hasil masing-masing nilai akurasi 80,137%, 78,0822%, 81,5060%, dan 83,5616%. terbaik di angka kombinasi *percentage split* 90% dengan presentase akurasi 83,5616%. Dan hasil evaluasi menggunakan *confusion matrix* menampilkan bahwa status konsumen yang terbanyak dan lebih unggul dari class karyawan dan class pelajar ialah menunjukkan Class Mahasiswa.

Kata kunci: prediksi; klasifikasi; naïve bayes; data mining; penjualan.

© 2020 Jurnal JOINTECS

Diterima Redaksi : 23-06-2020 | Selesai Revisi : 09-07-2020 | Diterbitkan Online : 30-09-2020

1. Pendahuluan

Celana Bahan adalah kebutuhan pokok yang dibutuhkan oleh manusia, bahkan semakin seiring berkembangnya zaman, celana bahan salah satu jenis simbol bahwa orang tersebut mempunyai kedudukan atau tidak. Salah satu fungsi celana adalah untuk menjaga penampilan agar tetap nyaman saat berinteraksi dengan orang lain. Pada umumnya celana bahan diwaktu lalu ialah digemari para karyawan dan identik dengan lanjut usia. Dari perkembangan zaman celana bahan memiliki berbagai macam model seperti Slimfit, Begi dan Formal. Mulai dizaman ini celana bahan sangat digemari namun masih tidak memiliki kepastian apakah dikalangan pelajar atau biasa disebut anak muda, dan juga kalangan mahasiswa atau masih seperti dulu masih kalangan karyawan yang lanjut usia yang menyukainya.

Dengan cara yang harus dilakukan ialah memprediksi Status konsumen pada produk celana bahan untuk mempermudah penjual berbasis UMKM (Usaha Mikro Kecil Menengah). Untuk menentukan stock penjualan pada status konsumen yang paling sering melakukan pembelian dan meningkatkan produksi penjualan [1]. Dengan memanfaatkan teknologi yang ada, adanya data mining bisa memanfaatkan untuk melakukan pengolahan data dan informasi yang didapatkan saat penjualan [2].

Data mining merupakan penggalian atau pengelolaan data. Dan juga mengartikan data mining menggali dan mengurai pengetahuan dari jumlah data yang ada. Data dan informasi memiliki nilai dan pengertian yang berbeda beda. Data dikatakan sebagai bahan mentah dari sebuah informasi tetapi informasi sendiri ialah data yang sudah dikelola dengan baik sehingga memiliki nilai tambah serta dapat digunakan dalam pengambilan keputusan [3].

Pada penelitian yang pernah dilakukan metode *Naive Bayes* juga digunakan dalam memprediksi Kelulusan dengan studi kasus data mahasiswa baru perguruan tinggi dengan memanfaatkan keakuratan nilai presentase pada dataset penerimaan mahasiswa baru menggunakan *Naive Bayes* [4]. Metode *Naive Bayes* ini dinilai baik dalam melakukan klasifikasi pada prediksi profesi berdasarkan *Skill Job Seeker* atau disebut calon pelamar pekerjaan, dibandingkan menggunakan metode lain [5]. Dibandingkan menggunakan metode *Algoritma C4.5*, merupakan pengembangan dari *Algoritma ID3* memiliki ketidakstabilan dalam melakukan klasifikasi. Dengan dibuktikannya hasil uji perbandingan pada prediksi penawaran produk [6]. Menghasilkan rata-rata akurasi yang didapatkan setelah tiga kasus uji data, metode *Naive Bayes* dengan nilai 57% dan *Algoritma C4.5* diangka 47,7% [6].

Dan juga penelitian terdahulu yang berhubungan dengan penjualan bahwa prediksi pada potensi pemasaran produk baru menggunakan metode *naive bayes* dan *regresi linear* merupakan penjualan produk dibidangan

makanan dan minuman [7]. Serta pada prediksi rating pada rievew produk kecantikan didapatkan hasil akurasi sebesar 87% [1]. Dan juga memanfaatkan *Naive Bayes* sebagai prediksi untuk rating penjualan buku dengan hasil akurasi yang di dapatkan sebesar 74,60% cukup baik dalam memprediksi [8]. Selain itu pada prediksi perilaku pola pembelian terhadap waktu transaksi dengan total 1768 data uji mencapai nilai akurasi 97% merupakan nilai yang sangat baik untuk metode *Naive Bayes* [9].

Untuk ukuran klasifikasi termasuk performasi terhadap tahapan hasil evaluasi menggunakan metode *Naive Bayes* [10]. Klasifikasi ini menentukan *record* data baru kesalah satu dari beberapa kategori *Class* yang telah didefinisikan sebelumnya. Biasanya hal ini disebut juga dengan *supervised learning* [11].

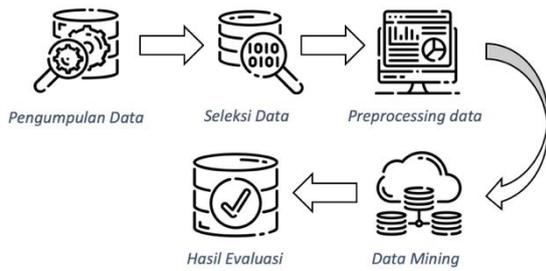
Pada penelitian ini prediksi status konsum produk celana, klasifikasi yang digunakan metode *Naive Bayes*. *Naive Bayes* suatu metode klasifikasi menggunakan metode probalitas dan statistic yang memanfaatkan *Confusion Matrix* [12]. setiap kelas keputusan, menghitung probabilitas dengan syarat untuk kelas keputusan ialah benar, mengasumsikan bahwa atribut obyek adalah sebuah independent [13].

Dengan tujuan penelitian ini ialah untuk mengetahui prediksi Status konsumen dengan model produk celana yang paling banyak diminati saat pembelian menggunakan *Naive Bayes*. Dengan rumusan masalah yang ada, bagaimana melakukan prediksi untuk mengkategorikan tipe status konsumen dengan model produk celana karena status konsumen tidak bisa dilihat hanya karena umur, bisa jadi diumur remaja dengan status pekerja atau diumur tua dengan status mahasiswa. Dengan batasan masalah yang dimiliki hanya membahas prediksi status konsumen pada produk celana, data yang gunakan data penjualan lengkap dari januari hingga desember 2019 dengan pengolahan data menggunakan *Naive Bayes*.

Dengan demikian penelitian ini akan mencoba mengimplementasikan data mining dengan algoritma *Naive Bayes* untuk prediksi Status Konsumen pada penjual berbasis UMKM (Usaha Mikro Kecil Menengah). Dengan memanfaatkan kinerja *Confusion Matrix* semoga bermanfaat untuk menentukan memprediksi tipe status konsumen pada produk celana bahan yang paling banyak diminati saat pembelian kedepannya khususnya bagi penjual atau pelaku usaha lainnya yang berbasis UMKM (Usaha Mikro Kecil Menengah).

2. Metode Penelitian

Langkah–langkah pada penelitian ini ialah ada beberapa langkah, yaitu: pengumpulan data, seleksi data, *preprocessing* data, data mining, hasil evaluasi. Dengan alur proses tahapan penelitian ini untuk mempermudah dan memahami setiap langka-langkah saat penelitian terjadi. Untuk lebih tepat dan jelas bisa dilihat pada Gambar 1.



Gambar 1. Tahap penelitian

2.1. Pengumpulan data

Pada tahap pengumpulan data ini dilakukan dengan cara wawancara pada pemilik penjual berbasis UMKM. lalu hasil penjualan pada bulan januari hingga desember tahun 2019 di *import* ke *excel*. Dengan memiliki 10 variabel dan setiap variabel memiliki beberapa *attribute*. Untuk melakukan ini supaya mempermudah saat proses klasifikasi prediksi dan mempermudah bagi penjual untuk mengetahui prediksi kalangan status mana pada pembeli. Pada Tabel 1 contoh beberapa hasil data penginputan wawancara ke *excel* pada beberapa data dibulan januari 2019.

Tabel 1. Contoh beberapa data bulan januari

Januari						
Model	Ukuran	Usia	Warna	Berat Badan	Jenis Kain	Kuant
Slimfit	37	25	Hitam	80-85	Drell	1
Slimfit	28	34	Cream	50-55	Drell	1
Slimfit	28	34	Abu-abu	50-55	Drell	1
Slimfit	30	32	Hitam	50-55	Drell	1
Slimfit	30	32	Hitam	50-55	Drell	1

Pada Tabel 1 menampilkan beberapa contoh hasil wawancara yang sudah di *import* ke *excel* pada bulan Januari dengan data sampai pada bulan desember ditahun 2019. Untuk setiap bulan memiliki jumlah data yang berbeda beda, diantaranya januari 56 data, februari 48 data, Maret 53 data, April 39 data, Mei 105 data, Juni 46 data, Juli 60 data, Agustus 52 data, September 91 data, Oktober 60 data, November 54 data, dan Desember 67 data dengan semua total pada tahun 2019 berjumlah 731 data.

2.2. Seleksi Data

Untuk mendapatkan hasil yang sempurna dan *spesifik* serta akurat. Dilakukan pemangkasan dan mengambil beberapa variabel agar memenuhi kriteria saat waktu proses klasifikasi [14]. Dengan awal mempunyai 10 variabel menjadi 5 variabel yang digunakan yaitu model, warna, usia, jenis kain dan status. Data yang akan digunakan mulai januari hingga desember tahun 2019 sebanyak 731 data.

2.3. Preprocessing data

Data yang telah didapatkan harus melakukan proses perbaikan data. Tujuan *preprocessing data* merupakan

pembersihan, penambahan, dan menyusun data menjadi terstruktur sesuai kebutuhan [15]. Data awal yang mempunyai 10 *variable* menjadi 5 *variable*, karena pengurangan *variable* tidak diperlukan agar proses pengujian data lebih mudah. sebelum tahap uji *preprocessing*, data tidak diketahui masing-masing *variable* memiliki berapa *attribute* dan beberapa *attribute* memiliki masing-masing jumlah data. Berikut beberapa contoh data proses perubahan data pada Tabel 2 dan Tabel 3.

Tabel 2. Data awal

Model	Ukuran	Usia	Warna	Berat Badan	Jenis Kain	Kuant
Slimfit	37	25	Hitam	80-85	Drell	1
Slimfit	28	34	Cream	50-55	Drell	1
Slimfit	28	34	Abu-abu	50-55	Drell	1
Slimfit	30	32	Hitam	50-55	Drell	1
Slimfit	30	32	Hitam	50-55	Drell	1

Pada Tabel 2 menampilkan beberapa data awal dan beberapa variabel dengan total yang dimiliki 10 variabel diantaranya Produk, Model, Ukuran, Kuantitas, Harga, Warna, Usia, Berat Badan, Jenis Kain, dan Status. Sebelum melakukan proses pengujian data awal merupakan data pengelompokan berdasarkan hasil penjualan perbulan ditahun 2019 lalu digabungkan menjadi 1 data kelompok. Berikut data yang digunakan saat proses pengujian dengan pengurangan beberapa variabel pada Tabel 3.

Tabel 3. Data digunakan

Model	Warna	Usia	Jenis Kain	Status
Slimfit	Hitam	25	Drell	Mahasiswa
Slimfit	Cream	34	Drell	Karyawan
Slimfit	Abu-abu	34	Drell	Karyawan
Slimfit	Hitam	32	Drell	Karyawan
Slimfit	Hitam	32	Drell	Karyawan

Untuk Tabel 3 diatas merupakan data yang digunakan untuk tahap pengujian data yang mempunyai 5 variabel diantaranya Model, Warna, Usia, Jenis Kain, dan Status.

2.4. Data Mining

Dalam metode pengujian ini menggunakan *Naïve Bayes*. *Naïve Bayes* ialah pengklasifikasian menggunakan metode metode probabilitas dan statistik yang dikemukakan oleh ilmuan inggris Thomas Bayes [16]. Dimana prediksi peluang dimasa depan berdasarkan pengalaman dimasa sebelumnya hingga dikenal dengan *Teorema Bayes* [16]. Rumus 1 umum *Teorema Bayes* [17].

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)} \quad (1)$$

Keterangan rumus 1 ada pada penjelasan ini. *x* merupakan data dengan *class* yang belum diketahui dan *c* yaitu hipotesis dari data *x* yang merupakan suatu *class* spesifik. Sedangkan *P(c|x)* merupakan probabilitas hipotesis *c* yang berdasarkan kondisi *x* (*posterior*

probability) lalu $P(c)$ probabilitas hipotesis c (*prior probability*) [18]. Untuk $P(x|c)$ yaitu probabilitas x berdasarkan kondisi hipotesis c dan $P(x)$ yaitu probabilitas dari x [10].

2.5. Hasil Evaluasi

Dalam hasil ini yang ditampilkan dalam bentuk mudah dimengerti oleh pihak pembaca. Dengan ini mencakup pola dan informasi yang ditemukan, menggunakan *Confusion Matrix* untuk mengevaluasi model klasifikasi yang digunakan memperkirakan objek yang benar dan yang salah [6]. Dengan hasil prediksi akan dibandingkan dengan kelas yang asli dari data tersebut. *Confusion Matrix* mengevaluasi kinerja model berdasarkan pada kemampuan prediksi suatu model [19]. Akurasi merupakan parameter yang mengukur ketepatan aturan klasifikasi yang dihasilkan dalam mengklasifikasi *Data Test Set* berdasarkan *attribute* yang ada dalam *Classnya* [6]. Gambar 2 *Confusion Matrix*.

		Kelas Sebenarnya	
		Yes	No
Kelas Yang diprediksi	Yes	True Positive	False Positive
	No	False Negative	True Negative

Gambar 2. *Confusion Matrix*

Penjelasan pada Gambar 2 ialah *True Positive* dengan jumlah data yang diklasifikasi *Yes* dan diprediksi *Yes* dengan benar. *False Positive* merupakan jumlah data yang diklasifikasikan *No* dan diprediksi *Yes*. *False Negative* yaitu jumlah data yang diklasifikasikan *Yes* dan diprediksi *No*. dan *True Negative* adalah jumlah data yang diklasifikasikan *No* dan diprediksi *No* dengan benar. Untuk hasil akurasi menggunakan pada rumus 2 [5].

$$Posterior = \frac{Prior \times Likelihood}{Evidence} \quad (2)$$

Penjelasan pada rumus 2 ialah *Posterior* merupakan peluang munculnya kelas sebelum masuknya sampel tersebut, seringkali disebut *prior*, dikali dengan peluang kemunculan karakteristik-karakteristik sampel pada kelas disebut juga *likelihood*, dibagi dengan peluang kemunculan karakteristik-karakteristik sampel secara global disebut juga *evidence*. Nilai *evidence* selalu tetap untuk setiap kelas pada satu sampel. Nilai dari *posterior* tersebut nantinya akan dibandingkan dengan nilai-nilai

posterior kelas lainnya untuk menentukan ke kelas apa suatu sampel akan diklasifikasikan [5].

3. Hasil dan Pembahasan

Hasil penelitian yang telah diujikan pada tahap ini menggunakan metode klasifikasi *Naive Bayes*. Sebelum tahap pengujian dilakukan tahap preprocessing. pada saat pengujian klasifikasi dilakukan 2 teknik yaitu menggunakan *Use Training Set* dan *Percentage Split*. Lalu yang terakhir melakukan tahap evaluasi *Confusion Matrix*.

3.1. Hasil Preprocessing

Data yang sebelumnya memiliki 10 variable menjadi 5 variable, diantaranya Model, Warna, Usia, Jenis Kain, Status yang berjumlah 731 Data. Hasil dari data masing-masing *Preprocessing* per *Attribute* bisa dilihat pada Tabel 4 hingga Tabel 8.

Tabel 4. Data Model

Model	Total
Slimfit	237
Formal	208
Begi	286

Dirujuk Tabel 4 data model menunjukkan hasil preprocessing pada variabel model celana kain, masing-masing *attribute* slimfit 237 data, formal 208 data, begi 286 data.

Tabel 5. Data Warna

Warna	Total
Hitam	506
Cream	51
Abu-abu	102
Biru dongker	72

Pada Tabel 5 data warna diatas setelah hasil *preprocessing* menunjukkan Hitam 506 data, Cream 51 data, Abu-Abu 102 data, Biru dongker 72 data.

Tabel 6. Data Usia

Usia	Total
15 – 19,364	207
19,364 – 23,727	182
23,727 – 28,091	134
28,091 – 32,455	46
32,455 – 36,818	48
36,818 – 41,182	47
41,182 – 45,545	2
45,545 – 49,909	18
49,909 – 54,273	4
54,273 – 58,636	20
58,636 – 63	23

Untuk Tabel 6 data usia menunjukkan hasil *preprocessing* dengan rentan usia 15 tahun hingga 63 tahun diantaranya, usia 15-19 207 data, 19-23 182 data, 23-28 134 data, 28-32 46 data, 32-36 48 data, 36-41 47 data, 41-45 2 data, 45-49 18 data, 49-54 4 data, 54-58 20

data, 58-63 2 data. Lalu pada Tabel 7 dibawah ini menampilkan data pada variabel jenis kain dengan memiliki 5 *attribute* yang masing-masing *attribute* memiliki jumlah data yang berbeda.

Tabel 7. Data Jenis Kain

Jenis Kain	Total
Drell	105
Teflon	445
Belini	54
Wool	54
High Twist	73

Untuk Tabel 7 menunjukkan data hasil *preprocessing* dengan 5 jenis kain diantaranya, Drell 105 data, Teflon 445 data, Belini 54 data, Wool 54 data, High Twist 73 data.

Tabel 8. Data Status

Status	Total
Mahasiswa	283
Karyawan	289
Pelajar	159

Pada Tabel 8 Data menunjukkan hasil *preprocessing* dengan hasil Mahasiswa 283 data, Karyawan 289 data, Pelajar 159 data.

3.2. Tahap Hasil Pengujian

Dalam hasil pengujian ini dilakukan langkah awal yaitu *Training* data menggunakan *Use Training Set* dengan total 731 data. Untuk hasil proses *Training* bisa dilihat pada Tabel 9.

Tabel 9. Hasil *Use Training Set*

<i>Correctly Classified Instances</i>	580	79,3434 %
<i>Incorrectly Classified Instances</i>	151	20,6565 %
<i>Kappa Statistic</i>	0,6805	
<i>Mean absolute error</i>	0,1928	
<i>Root mean Squared error</i>	0,3211	
<i>Relative absolute error</i>	44,7278 %	
<i>Root relative squared error</i>	69,1628 %	
<i>Total Number of Instances</i>	731	

Berdasarkan Tabel 9 presentase menggunakan *Use Training Set* untuk *Correctly Classified Instance* sebesar 79,3434% dan untuk persentase *Incorrectly Classified Instance* sebesar 20,6565%. Dimana 580 data berhasil diklasifikasikan dengan benar dan 151 data tidak berhasil diklasifikasikan dengan benar dari total 731 data.

Lalu pengujian data *testing* menggunakan *Percentage Split*. hasil yang dilakukan ini tentu berbeda beda, karena *Percentage Split* dilakukan beberapa kombinasi diantaranya 60%, 70%, 80%, dan 90%. Kombinasi ini dilakukan dengan memasukkan data untuk pelatihan, dan sisanya untuk data uji (*Testing*). Kombinasi diantara

60%, 70%, 80%, dan 90% yang sudah dilakukan untuk hasil masing-masing bisa dilihat pada Tabel 10.

Tabel 10. Hasil Kombinasi *Percentage Split*

<i>Percentage Split (%)</i>	Hasil
60	80,137 %
70	78,0822 %
80	81,5068 %
90	83,5616 %

Pada Tabel 10 Menampilkan hasil dari beberapa kombinasi *Percentage Split* diantaranya 60% dengan hasil 80,137%, 70% dengan hasil 78,0822%, 80% dengan hasil 81,5068% dan 90 dengan hasil 83,5616%. Untuk masing-masing detail hasil *Percentage split* pada Tabel 11 sampai Tabel 14.

Tabel 11. Hasil *Percentage Split* 60%

<i>Correctly Classified Instances</i>	234	80,137 %
<i>Incorrectly Classified Instances</i>	58	19,863 %
<i>Kappa Statistic</i>	0,6937	
<i>Mean absolute error</i>	0,1951	
<i>Root mean Squared error</i>	0,3249	
<i>Relative absolute error</i>	45,2362 %	
<i>Root relative squared error</i>	69,947 %	
<i>Total Number of Instances</i>	292	

Dirujuk Tabel 11 presentase menggunakan *Percentage Split* 60% ialah *Correctly Classified Instance* sebesar 80,137% dan untuk *Incorrectly Classified Instance* sebesar 19,863%. Yang dimana 234 data berhasil diklasifikasikan dengan benar dan 58 data tidak berhasil diklasifikasikan dengan benar dari data *Testing* 292 data.

Tabel 12. Hasil *Percentage Split* 70%

<i>Correctly Classified Instances</i>	171	78,0822 %
<i>Incorrectly Classified Instances</i>	48	21,9178 %
<i>Kappa Statistic</i>	0,6681	
<i>Mean absolute error</i>	0,1926	
<i>Root mean Squared error</i>	0,3322	
<i>Relative absolute error</i>	44,6182 %	
<i>Root relative squared error</i>	71,4878 %	
<i>Total Number of Instances</i>	219	

Tabel 12 presentase menggunakan *Percentage Split* 70% ialah *Correctly Classified Instance* sebesar 78,0822% dan untuk *Incorrectly Classified Instance* sebesar 21,9178%. Yang dimana 171 data berhasil diklasifikasikan dengan benar dan 48 data tidak berhasil diklasifikasikan dengan benar dari data *Testing* 219 data.

Tabel 13. Hasil *Percentage Split* 80%

<i>Correctly Classified Instances</i>	119	81,5068 %
<i>Incorrectly Classified Instances</i>	27	18,4932 %
<i>Kappa Statistic</i>	0,7066	
<i>Mean absolute error</i>	0,1804	
<i>Root mean Squared error</i>	0,3113	
<i>Relative absolute error</i>	42,1901 %	
<i>Root relative squared error</i>	67,8009 %	
<i>Total Number of Instances</i>	146	

Pada Tabel 13 presentase menggunakan *Percentage Split* dikombinasi 80% untuk *Correctly Classified Instance* sebesar 81,5068% dan untuk *Incorrectly Classified Instance* sebesar 18,4932%. Yang dimana 119 data berhasil diklasifikasikan dengan benar dan 27 data tidak berhasil diklasifikasikan dengan benar dari data *Testing* 146 data.

Tabel 14. Hasil *Percentage Split* 90%

<i>Correctly Classified Instances</i>	61	83,5616 %
<i>Incorrectly Classified Instances</i>	12	16,4384 %
<i>Kappa Statistic</i>	0,739	
<i>Mean absolute error</i>	0,1769	
<i>Root mean Squared error</i>	0,2948	
<i>Relative absolute error</i>	41,5409 %	
<i>Root relative squared error</i>	64,3805 %	
<i>Total Number of Instances</i>	73	

Untuk Tabel 14 presentase menggunakan *Percentage Split* dikombinasi 90% untuk *Correctly Classified Instance* sebesar 83,5616% dan untuk *Incorrectly Classified Instance* sebesar 16,4384%. Yang dimana 61 data berhasil diklasifikasikan dengan benar dan 12 data tidak berhasil diklasifikasikan dengan benar dari data *Testing* 73 data. Lalu yang terakhir yaitu tahap Hasil evaluasi yang menggunakan teknik *Confusion Matrix* yang menggunakan beberapa kombinasi dari 60% hingga 90%.

3.3. Hasil Evaluasi

Hasil evaluasi pengujian mengukur keakuratan hasil yang diuji, untuk akurasi itu sendiri ialah bisa dikatakan tingkat kedekatan antara nilai prediksi dan nilai aktual. Pengukuran akurasi pada tahap evaluasi dilakukan menggunakan *Confusion Matrix*. Pada Tabel 15 hingga 18 menampilkan hasil akurasi pada *Percentage Split* dikombinasi 60% hingga 90%.

Tabel 15. Hasil *Confusion Matrix* 60%

60 %	Prediksi		
	Mahasiswa	Karyawan	Pelajar
Class			
Mahasiswa	107	3	1
Karyawan	48	68	1
Pelajar	5	0	59

Pada Tabel 15 hasil *Confusion Matrix* terlihat *Class* Mahasiswa yang diprediksi tepat sebagai Mahasiswa sebanyak 107 data, 4 data tidak tepat diprediksi sebagai *Class Mahasiswa* karena data tersebut di prediksi di *Class Karyawan* 3 data dan *Pelajar* 1 data. Lalu pada *Class Karyawan* yang diprediksi tepat sebagai Karyawan sebanyak 68 data, untuk 49 data tidak tepat diprediksi sebagai Karyawan karena data tersebut diprediksi sebagai *Class* Mahasiswa 48 data dan *Pelajar* 1 data. Kemudian *Class Pelajar* tepat yang diprediksi sebagai *Pelajar* sebanyak 59 data, untuk 5 data tidak tepat diprediksi sebagai *Class Pelajar* karena data itu diprediksi sebagai *Class* Mahasiswa.

Tabel 16. Hasil *Confusion Matrix* 70%

70 %	Prediksi		
	Mahasiswa	Karyawan	Pelajar
Class			
Mahasiswa	70	3	4
Karyawan	36	55	4
Pelajar	1	0	46

Pada Tabel 16 hasil *Confusion Matrix* terlihat *Class* Mahasiswa yang diprediksi tepat sebagai Mahasiswa sebanyak 70 data, 7 data tidak tepat diprediksi sebagai *Class Mahasiswa* karena data tersebut di prediksi di *Class Karyawan* 3 data dan *Class Pelajar* 4 data. Lalu pada *Class Karyawan* yang diprediksi tepat sebagai Karyawan sebanyak 55 data, untuk 40 data tidak tepat diprediksi sebagai Karyawan karena data tersebut diprediksi sebagai *Class* Mahasiswa 36 data dan 4 data *Class Pelajar*. Kemudian *Class Pelajar* tepat yang diprediksi sebagai *Pelajar* sebanyak 46 data, untuk 1 data tidak tepat diprediksi sebagai *Class Pelajar* karena data itu diprediksi sebagai *Class* Mahasiswa.

Tabel 17. Hasil *Confusion Matrix* 80%

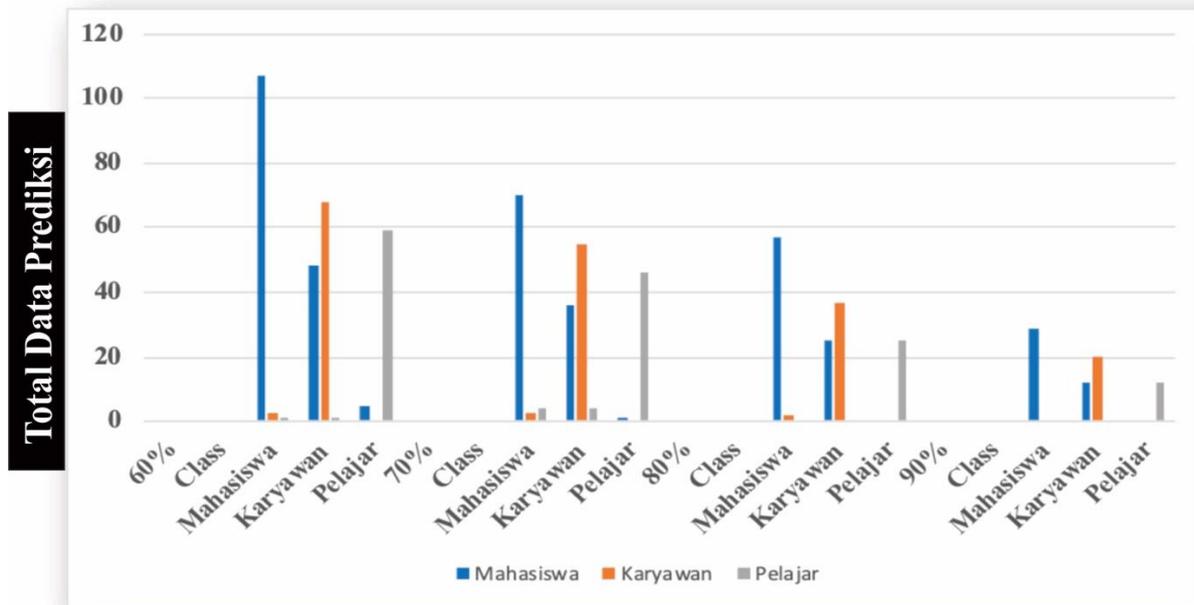
80 %	Prediksi		
	Mahasiswa	Karyawan	Pelajar
Class			
Mahasiswa	57	2	0
Karyawan	25	37	0
Pelajar	0	0	25

Pada Tabel 17 hasil *Confusion Matrix* terlihat *Class* Mahasiswa yang diprediksi tepat sebagai Mahasiswa sebanyak 57 data, 2 data tidak tepat diprediksi sebagai *Class Mahasiswa* karena data tersebut di prediksi di *Class Karyawan*. Lalu pada *Class Karyawan* yang diprediksi tepat sebagai Karyawan sebanyak 37 data, untuk 25 data tidak tepat diprediksi sebagai Karyawan karena data tersebut diprediksi sebagai *Class* Mahasiswa. Kemudian *Class Pelajar* tepat yang diprediksi sebagai *Pelajar* sebanyak 25 data.

Tabel 18. Hasil *Confusion Matrix* 90%

90 %	Prediksi		
	Mahasiswa	Karyawan	Pelajar
Class			
Mahasiswa	29	0	0
Karyawan	12	20	0
Pelajar	0	0	12

Pada Tabel 18 hasil *Confusion Matrix* terlihat *Class* Mahasiswa yang diprediksi tepat sebagai Mahasiswa sebanyak 29 data. Lalu pada *Class Karyawan* yang diprediksi tepat sebagai Karyawan sebanyak 20 data. Dan 12 data tidak tepat diprediksi sebagai Karyawan karena data tersebut diprediksi sebagai *Class* Mahasiswa. Kemudian *Class Pelajar* tepat yang diprediksi sebagai *Pelajar* sebanyak 12 data. Untuk lebih mudah memahami hasil yang sudah dilakukan maka ditampilkan visualisasi hasil dalam bentuk diagram pada Gambar 3.



Gambar 3. Diagram Hasil Prediksi

Dirujuk dari gambar 3 menampilkan diagram hasil prediksi menggunakan teknik beberapa kombinasi *Confusion Matrix* 60% sampai 90% diantaranya berwarna biru merupakan *class* Mahasiswa, berwarna orange pada *class* Karyawan, dan berwarna abu-abu merupakan *class* Pelajar. Pada sumbu Y merupakan Total data yang diprediksi dan untuk sumbu X merupakan kombinasi *confusion matrix* 60% hingga 90% yang masing-masing menampilkan 3 *class*. Untuk *class* mahasiswa berwarna biru yang selalu tertinggi pada setiap kombinasi *confusion Matrix* karena *class* mahasiswa merupakan *class* dengan jumlah terbanyak yang selalu diprediksi dengan benar. Pada *class* karyawan yang berwarna orange memiliki jumlah yang diprediksi akurat lebih sedikit dari pada *class* mahasiswa karena pada *class* tersebut salah memprediksi yang tertuju pada *class* mahasiswa berwarna biru. Akan tetapi yang memiliki jumlah paling sedikit dimiliki oleh *class* pelajar yang berwarna abu-abu karena *class* tersebut selalu memprediksi dengan benar dan juga paling sedikit memprediksi tidak tepat pada *class*nya.

Hasil prediksi tersebut jika dalam kondisi normal diluar adanya kondisi luar biasa seperti terjadinya pandemi. Rupanya prediksi juga bisa meleset jika terjadi pandemi karena data tersebut diambil diluar dari adanya pandemi karena pada saat terjadi pandemi lapak ditutup untuk keamanan bersama dan dibuka kembali setelah pemberitahuan dari pemerintah daerah setempat dengan menerapkan protokol kesehatan.

4. Kesimpulan

Dalam kesimpulan ini menunjukkan hasil dari klasifikasi *Naïve Bayes* dengan menggunakan

Percentage Split beberapa kombinasi 60 hingga 90 persen menunjukkan bahwa *Percentage Split* 90% mempunyai nilai tertinggi diangka 83,56% merupakan yang terbaik. Untuk yang mempunyai nilai terendah di *Percentage Split* 70% yaitu 78,0822%, Dengan rata-rata nilai hasil akurasi diangka 80,82%. Memanfaatkan kinerja *Confusion Matrix* mengetahui prediksi setiap kombinasi *Percentage Split* mulai 60 hingga 90 persen bahwa *Class* Mahasiswa merupakan status konsumen yang selalu miliki jumlah data yang lebih unggul dibanding *Class* Karyawan dan *Class* Pelajar, Dengan ini Prediksi Status Konsumen Produk Celana yang paling banyak pada kalangan Mahasiswa. Semoga bermanfaat bagi yang memerlukan, khususnya bagi para penjual dan pelaku usaha lainnya dalam kategori UMKM (Usaha Mikro Kecil Menengah).

Daftar Pustaka

- [1] F. Rosi, M. A. Fauzi, and R. S. Perdana, "Prediksi Rating Pada Review Produk Kecantikan Menggunakan Metode Naïve Bayes dan Categorical Proportional Difference (CPD)," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer (J-PTIIK) Universitas Brawijaya*, 2018. .
- [2] G. Schuh, J.-P. Prote, and P. Hünnekes, "Data mining methods for macro level process planning," *Procedia CIRP*, vol. 88, pp. 48–53, Jan. 2020.
- [3] S. F. Rodiyansyah, "Algoritma Apriori untuk Analisis Keranjang Belanja pada Data Transaksi Penjualan," *Infotech*, vol. 1, no. 1, pp. 36–39, 2015.

- [4] S. Syarli and A. A. Muin, "Metode Naive Bayes Untuk Prediksi Kelulusan (Studi Kasus: Data Mahasiswa Baru Perguruan Tinggi)," *J. Ilmu Komput.*, vol. 2, no. 1, pp. 22–26, 2018.
- [5] D. Hastuti *et al.*, "Algoritma Naive Baiyes Untuk Prediksi Profesi Berdasarkan Skill Job Seeker," in *Seminar Nasional Riset Kuantitatif Terapan 2017*, 2017, vol. 1, no. April, pp. 63–67.
- [6] A. Khoerunnisa, B. Irawan, and M. R. Rumani, "Analisis dan implementasi perbandingan algoritma c.45 dengan naive bayes untuk prediksi penawaran produk," *E-Proceeding Eng.*, vol. 3, no. 3, pp. 5029–5035, 2016.
- [7] T. I. Andini, W. Witanti, and F. Renaldi, "Prediksi Potensi Pemasaran Produk Baru dengan Metode Naive Bayes Classifier dan Regresi Linear," *Semin. Nas. Apl. Teknol. Inf.*, pp. 27–32, 2016.
- [8] Y. S. Nugroho, "Penerapan Data Mining Untuk Prediksi Rating Penjualan Buku Menggunakan Metode Naive Bayes Sulastris * , Yusuf Sulisty Nugroho," *Duta.com*, vol. 12, no. 2, pp. 57–72, Apr. 2017.
- [9] S. Kom, E. Dewi, S. Mulyani, S. Kom, and I. R. Nurhasanah, "Penerapan Data Mining Classification Untuk Prediksi Perilaku Pola Pembelian Terhadap Waktu Transaksi Menggunakan Metode Naive Bayes," *Konf. Nas. Sist. dan Inform.*, pp. 9–10, 2015.
- [10] F. Xu, Z. Pan, and R. Xia, "E-commerce product review sentiment classification based on a naive Bayes continuous learning framework," *Inf. Process. Manag.*, vol. 57, no. 5, p. 102221, Sep. 2020.
- [11] C. Catur, "Prediksi Kegagalan Siswa Dalam Data Mining Dengan," vol. 3, no. 1, pp. 42–46, 2019.
- [12] J. Singh, S. Bagga, and R. Kaur, "Software-based Prediction of Liver Disease with Feature Selection and Classification Techniques," in *Procedia Computer Science*, 2020, vol. 167, pp. 1970–1980.
- [13] S. Alfarisi, "Sistem Prediksi Penjualan Gamis Toko QITAZ Menggunakan Metode Single Exponential Smoothing," *JABE (Journal Appl. Bus. Econ.)*, vol. 4, no. 1, pp. 80–95, 2017.
- [14] W. Hadi, Q. A. Al-Radaideh, and S. Alhawari, "Integrating associative rule-based classification with Naive Bayes for text classification," *Appl. Soft Comput. J.*, vol. 69, pp. 344–356, Aug. 2018.
- [15] R. Andean, S. Fendy, and A. Nugroho, "Klasterisasi Pengendalian Persediaan Aki," *J. Inf. Technol. Comput. Sci.*, vol. 4, no. 1, pp. 5–11, 2019.
- [16] B. As'ad, "Prediksi Kehadiran Menggunakan Metode Klasifikasi Naive Bayes, One-r, Decision Tree," *J. Penelit. Komun. dan Opini Publik*, vol. 20, no. 1, pp. 1–10, 2016.
- [17] D. Laia, E. Buulolo, and M. J. F. Sirait, "Implementasi Data Mining Untuk Memprediksi Pemesanan Driver Go-Jek Online Dengan Menggunakan Metode Naive Bayes (Studi Kasus: Pt. Go-Jek Indonesia)," in *KOMIK (Konferensi Nasional Teknologi Informasi dan Komputer)*, 2018, vol. 2, no. 1, pp. 434–439.
- [18] Y. Ko, "How to use negative class information for Naive Bayes classification," *Inf. Process. Manag.*, vol. 53, no. 6, pp. 1255–1268, Nov. 2017.
- [19] H. M and S. M.N, "A Review on Evaluation Metrics for Data Classification Evaluations," *Int. J. Data Min. Knowl. Manag. Process*, vol. 5, no. 2, pp. 01–11, 2015.