

Application of Data Mining With Naïve Bayes Classifier Algorithm to Predict Customer Character: Penerapan Data Mining Dengan Algoritma Naïve Bayes Classifier untuk Memprediksi Karakter Pelanggan

Anisa Yuli Eka Pratiwi
Arif Senja Fitriani

Universitas Muhammadiyah Sidoarjo
Universitas Muhammadiyah Sidoarjo

This research is based on late payment data from UD Delta customers which affect all production processes that occur in UD Delta. The purpose of this study is to predict the character of the customer as a basis for analyzing the delay in the payment process that hinders the company's production. The datasets tested were 30% of the 600 data, which were divided into 1 id, 7 attributes and for the prediction classes 'No Problem', 'Slightly Problematic', and 'Very Problematic'. This study collects data by means of non-participant observation, namely data on UD Delta's bookkeeping. The process stage is processed using the python programming language with classification methods and naive bayes algorithms. The results obtained from this study are in the form of calculations that state the number of customers grouped into 3 categories, namely with correct predictions as much as 51%, incorrect predictions as much as 49%. From the results obtained, it will be taken into consideration by the company in processing the existing processes within the company as a step to develop the business carried out so that it is more optimal.

Pendahuluan

Perusahaan adalah suatu unit kegiatan produksi yang mengolah sumber-sumber ekonomi untuk menyediakan barang dan jasa bagi masyarakat dengan tujuan untuk memperoleh keuntungan agar dapat memuaskan kebutuhan masyarakat[1]. Perusahaan juga didefinisikan sebagai suatu organisasi produksi yang menggunakan dan mengkoordinir sumber-sumber ekonomi untuk memuaskan kebutuhan dengan cara yang menguntungkan. Dari definisi tersebut ada 5 unsur yang penting, yaitu organisasi, produksi, sumber-sumber ekonomi atau factor-faktor produksi, kebutuhan dan cara menguntungkannya, cara yang menguntungkan tersebut diartikan sebagai cara yang diambil perusahaan harus memperhatikan prinsip efisiensi. Salah satu prinsip yang harus diperhatikan yaitu tujuan perusahaan [2].

Ditinjau dari penelitian yang dilakukan pada UD Delta, tujuan badan usaha ini berkonsentrasi pada kelangsungan hidup dan pertumbuhan perusahaan. Kedua hal tersebut bisa didapatkan pada semua proses jual beli pada perusahaan berjalan lancar dan efisien. Karenanya factor pelanggan sangat berpengaruh dalam proses jual beli yang dilaksanakan.

Pada kasus yang terjadi di UD Delta, usaha dagang yang selama ini berjalan terhambat oleh pelanggan yang masih memiliki catatan pembayaran yang belum lunas pada bagian administrasi pembayaran barang. Tercatat UD Delta memiliki sebanyak 24 pelanggan yang membeli bebrbagai

macam produk dari 40 jenis produk yang ditawarkan oleh perusahaan. Dan sebagian dari pelanggan tersebut masih melakukan rutinitas pengambilan produk secara bertahap dimana catatan piutang yang masih tercatat menjadi beban perusahaan. Dikarenakan banyaknya piutang dari pelanggan yang belum terbayarkan tersebut mempengaruhi proses produksi dari UD Delta. Dari permasalahan tersebut dilakukan penelitian ini dengan tujuan dapat membantu UD Delta dalam menganalisa keterlambatan proses pembayaran pelanggan kepada perusahaan untuk memaksimalkan keuntungan yang didapatkan perusahaan dengan metode Naïve Bayes Classifier, menggunakan bahasa pemrograman Python berbasis WEB [3-7].

Metode Penelitian

Dalam Penelitian, pencarian serta pengumpulan data dilakukan terlebih dahulu. Kegiatan pencarian dan pengumpulan data yang dilakukan penulis untuk mendukung penelitian yaitu dengan melakukan Studi Literatur dimana pengumpulan data yang diperoleh melalui berbagai artikel yang didapat dari jurnal dan buku-buku referensi yang mendukung penyelesaian dari permasalahan yang ada. Setelah itu melakukan pengamatan (non participant Observation) [8]. Pada tahap observasi, peneliti menggunakan metode non participant observation. Kegiatan yang dilakukan penulis adalah menginputkan data berupa hardcopy pada pembukuan UD Delta kedalam Soft Copy dengan ekstensi .csv. Peneliti mengumpulkan data sebanyak 600 data dari record data transaksi yang dilakukan 24 pelanggan yang melakukan transaksi dengan UD Delta.

Perancangan Diagram Klasifikasi

Gambar 1. Diagram Klasifikasi

Pada Gambar 1. Diagram Klasifikasi dapat dijelaskan alur mengklasifikasi sebagai solusi dari permasalahan yang diangkat oleh penulis. Dimulai dari menginputkan data dari pembukuan kedalam komputer lalu data tersebut akan melalui proses pre-processing [9] yang terdiri dari data cleaning, data integrasi lalu menentukan atribut dan fitur dan melalui proses transformasi [10]. Setelah melalui proses tersebut akan menjadi Dataset [11]. Dataset inilah yang akan dibagi menjadi 2 yaitu Data Training dan Data Testing. Setelah data tersebut didapatkan, selanjutnya menentukan probabilitas dari Data Training untuk menghitung nilai probabilitasnya [12]. Proses selanjutnya menghitung nilai Bayes berdasarkan nilai Probabilitas yang sudah didapat lalu menerapkan Algoritma Naïve Bayes pada Data Testing [13-15]. Dari penerapan Algoritma Naïve Bayes pada Data Testing akan menghasilkan output berupa hasil keterangan bahwa pelanggan tersebut bermasalah atau tidak bermasalah.

Preprocessing Data

Dalam Preprocessing data, penulis melakukan beberapa tahap yaitu:

Dalam pengklasifikasian karakter pelanggan UD Delta, difokuskan pada data Transaksi pelanggan. Dari data transaksi pelanggan tersebut dilakukan proses feature extraction.

Proses ini menata ulang data dengan membuang data yang tidak dibutuhkan lalu memperbaiki data yang terdapat salah penulisan seperti tanda titik ('.'), tanda koma (',').

- Feature extraction
- Cleaning data.
- Feature Selection and Transformation (Seleksi Fitur dan Transformasi)

Menyeleksi Fitur dan Transformasi data dilakukan untuk memetakan fitur dengan mengganti fitur numerik ke dalam konseptual label, diantaranya pada beberapa fitur pada Tabel 1. Variabel dan Fitur. Pada Transformasi data ditentukan Variabel dan Fitur. Atribut (variable) merupakan parameter yang menyebabkan class/ label/target terjadi. Ada 7 atribut yang digunakan, yaitu

pelanggan, termin pembayaran, potongan, daerah, piutang, deadline pembayaran, volume transaksi dan dengan 1 id dan 1 goal yaitu status.

Dataset

Dataset adalah sekumpulan data dari data transaksi pelanggan yang telah direkap kedalam bentuk soft file dari pembukuan yang ada pada UD Delta. Pada penelitian ini menggunakan sebanyak 60 data sample secara acak yang sudah melalui tahap preprocessing dan transformasi. Data set tersebut akan dibuat untuk tahapan perhitungan manual dengan standart 70% Data Training dan 30% sebagai data Testing.

A. Variabel dan fitur

No	Variabel	Fitur	Diskripsi
1	Invoice	TID	ID Invoce
2	Pelanggan	Kategorikal	Nama
3	Termin Pembayaran	Kategorikal { Lunas, Belum Lunas }	Lunas = Terjadi Pembayaran dengan termin. Belum Lunas = Tidak terjadi pembayaran dengan termin
4	Potongan	Kategorikal { ada, tidak }	Ada = Tercatat memiliki potongan harga. Tidak = tidak ada catatan potongan harga.
5	Daerah	Kategorikal { GRK, PRM, LMG, PSR, SDA, SBY, MJK, MDR, LMJG, MLG, BJN, JBG }	GRK=Gresik, PRM=Prambon, LMG=Lamongan, PSR=Pasuruan, SDA=Sidoarjo, SBY=Surabaya, MJK=Mojokerto, MDR=Madura, LMJG=Lumajang, MLG=Malang, BJN=Bojonegoro, JBG=Jombang.
6	Piutang	Kategorikal { Ada, Tidak }	Ada = Tercatat memiliki hutang. Tidak = Tida ada catatan hutang.
7	Deadline Pembayaran	Kategorikal { Tepat, Sedang, Tidak Tepat }	Tepat = Tepat waktu pembayaran. Sedang = Pelunasan pembayaran transaksi kemaren bersamaan dengan pelunasan selanjutnya. Tidak Tepat = pelunasan pembayaran dilakukan 3 kali atau lebih.
8	Volume Transaksi	Kategorikal { Sedikit, Sedannng, Lumayan Banyak, Banyak. }	Sedikit=Pembayaran<500.00 0. Sedang=500.000<pembayaran<1jt. Lumayan Banyak = pembayaran 1jt<pembayaran<3jt. Banyak = pembayaran>3jt.
9	Status	kategorikal { Tidak Bermasalah, Sedikit Bermasalah, Sangat Bermasalah }	Kelas prediksi. Tidak Bermasalah=tidak tercatat memiliki hutang. Sedikit Bermasalah. Sedang tercatat memiliki hutang. Sangat bermasalah=Tercatat memiliki hutang.

Table 1. Variabel dan Fitur

Tabel 1. Variabel dan Fitur merupakan salah satu tahap preprocessing data yaitu feature Selection dan Transformation. Pada proses ini terjadi seleksi fitur dan tranformasi data yang dilakukan untuk memetakan fitur dengan mengganti fitur numerik ke dalam konseptual label. Pada Transformasi data ditentukan Variabel dan Fitur seperti pada Tabel 1. Variabel dan Fitur.

B. Data training dan testing

TID	Termin Pembayaran	Potongan	Daerah	Piutang	Deadline Pembayaran	Volume Transaksi	Status
1	Lunas	Ada	Lumajang	Tidak ada	Tepat	Banyak	Tidak Bermasalah
2	Tidak Lunas	Tidak	Bojonegoro	Ada	Sedang	Banyak	Sedikit Bermasalah
3	Lunas	Tidak	Mojokerto	Tidak ada	Tepat	Sedang	Tidak Bermasalah
4	Lunas	Tidak	Mojosari	Tidak ada	Tepat	Lumayan Banyak	Tidak Bermasalah
5	Lunas	Ada	Lumajang	Tidak ada	Tepat	Banyak	Tidak Bermasalah
6	Tidak Lunas	Tidak	Jombang	Ada	Sedang	Lumayan Banyak	Sedikit Bermasalah
7	Lunas	Tidak	Perambon	Tidak ada	Tepat	Sedikit	Tidak Bermasalah
8	Lunas	Tidak	Sidoarjo	Tidak ada	Tepat	Sedikit	Tidak Bermasalah
9	Tidak Lunas	Tidak	Jombang	Ada	Sedang	Banyak	Sedikit Bermasalah
10	Lunas	Tidak	Mojosari	Tidak ada	Tepat	Lumayan Banyak	Tidak Bermasalah
11	Lunas	Tidak	Madura	Tidak ada	Tepat	Banyak	Tidak Bermasalah
12	Lunas	Tidak	Mojokerto	Tidak ada	Tepat	Sedikit	Tidak Bermasalah
13	Tidak Lunas	Tidak	Probolinggo	Ada	Tidak Tepat	Banyak	Bermasalah
14	Tidak Lunas	Tidak	Bojonegoro	Ada	Sedang	Banyak	Sedikit Bermasalah
15	Lunas	Tidak	Malang	Tidak ada	Tepat	Sedikit	Tidak Bermasalah
16	Lunas	Tidak	Pasuruan	Tidak ada	Tepat	Sedikit	Tidak Bermasalah
17	Lunas	Tidak	Surabaya	Tidak ada	Tepat	Sedang	Tidak Bermasalah
18	Tidak Lunas	Tidak	Malang	Ada	Tidak Tepat	Banyak	Bermasalah
19	Lunas	Tidak	Madura	Tidak ada	Tepat	Banyak	Tidak Bermasalah
20	Lunas	Tidak	Madura	Tidak ada	Tepat	Banyak	Tidak Bermasalah
21	Lunas	Ada	Lumajang	Tidak ada	Tepat	Banyak	Tidak Bermasalah
22	Tidak Lunas	Tidak	Probolinggo	Ada	Sedang	Lumayan Banyak	Sedikit Bermasalah
23	Lunas	Tidak	Surabaya	Tidak ada	Tepat	Sedikit	Tidak Bermasalah
24	Lunas	Tidak	Probolinggo	Tidak ada	Tepat	Banyak	Tidak Bermasalah
25	Lunas	Tidak	Mojokerto	Tidak ada	Tepat	Sedikit	Tidak Bermasalah
26	Lunas	Tidak	Surabaya	Tidak ada	Tepat	Lumayan Banyak	Tidak Bermasalah

27	Lunas	Tidak	Madura	Tidak ada	Tepat	Banyak	Tidak Bermasalah
28	Lunas	Tidak	Gresik	Tidak ada	Tepat	Sedikit	Tidak Bermasalah
29	Lunas	Tidak	Sidoarjo	Tidak ada	Tepat	Sedikit	Tidak Bermasalah
30	Tidak Lunas	Tidak	Probolinggo	Ada	Sedang	Lumayan Banyak	Sedikit Bermasalah
31	Lunas	Tidak	Surabaya	Tidak ada	Tepat	Sedang	Tidak Bermasalah
32	Lunas	Tidak	Mojosari	Tidak ada	Tepat	Lumayan Banyak	Tidak Bermasalah
33	Lunas	Tidak	Madura	Tidak ada	Tepat	Banyak	Tidak Bermasalah
34	Lunas	Tidak	Pasuruan	Tidak ada	Tepat	Sedikit	Tidak Bermasalah
35	Lunas	Tidak	Pasuruan	Tidak ada	Tepat	Sedang	Tidak Bermasalah
36	Lunas	Tidak	Lamongan	Tidak ada	Tepat	Lumayan Banyak	Tidak Bermasalah
37	Lunas	Ada	Pasuruan	Tidak ada	Tepat	Banyak	Tidak Bermasalah
38	Lunas	Tidak	Perambon	Tidak ada	Tepat	Sedikit	Tidak Bermasalah
39	Lunas	Tidak	Sidoarjo	Tidak ada	Tepat	Sedikit	Tidak Bermasalah
40	Lunas	Tidak	Madura	Tidak ada	Tepat	Sedikit	Tidak Bermasalah
41	Lunas	Tidak	Madura	Tidak ada	Tepat	Banyak	Tidak Bermasalah
42	Lunas	Tidak	Surabaya	Tidak ada	Tepat	Lumayan Banyak	Tidak Bermasalah

Table 2. *Data Training*

Tabel 2. Data Training merupakan data yang diambil acak dari data set yang berjumlah 60 data menjadi 42 data dimana data tersebut diambil sebanyak 70% untuk dijadikan Data Training. Data Training tersebut akan digunakan untuk menentukan nilai probabilitas pada setiap variable yang selanjutnya akan dilakukan proses perhitungan terhadap Data Testing atau Data Uji.

TID	Termin Pembayaran	Potongan	Daerah	Piutang	Deadline Pembayaran	Volume Transaksi	Status
1	Lunas	Tidak	Prambon	Tidak ada	Tepat	Sedikit	Tidak Bermasalah
2	Tidak Lunas	Tidak	Malang	Ada	Tidak Tepat	Banyak	Bermasalah
3	Lunas	Ada	Lumajang	Tidak ada	Tepat	Sedang	Tidak Bermasalah
4	Tidak Lunas	Tidak	Bojonegoro	Ada	Sedang	Sedikit	Sedikit Bermasalah
5	Lunas	Tidak	Pasuruan	Tidak ada	Tepat	Sedikit	Tidak Bermasalah
6	Lunas	Tidak	Malang	Tidak ada	Tepat	Sedikit	Tidak Bermasalah
7	Lunas	Tidak	Madura	Tidak ada	Tepat	Banyak	Tidak Bermasalah
8	Lunas	Tidak	Madura	Tidak ada	Tepat	Banyak	Tidak Bermasalah
9	Lunas	Tidak	Probolinggo	Tidak ada	Tepat	Sedikit	Tidak Bermasalah

10	Tidak Lunas	Tidak	Malang	Ada	Tidak Tepat	Banyak	Bermasalah
11	Lunas	Tidak	Gresik	Tidak ada	Tepat	Sedikit	Tidak Bermasalah
12	Lunas	Tidak	Lamongan	Tidak ada	Tepat	Lumayan Banyak	Tidak Bermasalah
13	Lunas	Tidak	Surabaya	Tidak ada	Tepat	Lumayan Banyak	Tidak Bermasalah
14	Lunas	Tidak	Mojosari	Tidak ada	Tepat	Banyak	Tidak Bermasalah
15	Lunas	Tidak	Sidoarjo	Tidak ada	Tepat	Sedikit	Tidak Bermasalah
16	Lunas	Tidak	Surabaya	Tidak ada	Tepat	Sedikit	Tidak Bermasalah
17	Lunas	Tidak	Pasuruan	Tidak ada	Tepat	Sedikit	Tidak Bermasalah
18	Lunas	Tidak	Mojokerto	Tidak ada	Tepat	Sedang	Tidak Bermasalah

Table 3. *Data Testing*

Tabel 3. Data Testing merupakan data yang diambil secara acak dari data set yang berjumlah 60 data menjadi 18 data dimana data tersebut diambil sebanyak 30% dari Data Set untuk dijadikan Data Testing atau data Uji.

Hasil dan Pembahasan

1. Proses Perhitungan Manual

Tahap Proses merupakan perhitungan manual untuk menentukan probabilitas dari data training dan menerapkan Algoritma Naïve Bayes pada Data Testing yang sudah ditentukan.

Untuk menentukan Probabilitas data target atau Class harus diketahui jumlah data terlebih dahulu.

Total Data Training= 42

Jumlah Status Tidak Bermasalah= 34

Jumlah Status Sedikit Bermasalah= 6

Jumlah Status Sangat Bermasalah= 2

Probabilitas dengan rumus berikut:

Status Tidak Bermasalah:

Status Sedikit Bermasalah:

Status Sangat Bermasalah:

Lalu dilanjutkan menghitung probabilitas tiap variable:

Tidak Bermasalah	Sedikit Bermasalah	Sangat Bermasalah
P(tidak bermasalah/sedikit bermasalah /sangat bermasalah)	0.80952381	0.142857143

	1	
P(pembayaran)	Tidak Bermasalah	Sedikit Bermasalah
Lunas	1	0
Tidak Lunas	0	1
	1	1
P(potongan)	Tidak Bermasalah	Sedikit Bermasalah
Ada	0.117647059	0
Tidak	0.882352941	1
	1	1
P(Daerah)	Tidak Bermasalah	Sedikit Bermasalah
Lumajang	0.088235294	0
Bojonegoro	0	0.333333333
Mojokerto	0.088235294	0
Mojosari	0.088235294	0
Jombang	0	0.333333333
Perambon	0.058823529	0
Sidoarjo	0.088235294	0

Madura	0.205882353	0
Probolinggo	0.029411765	0.333333333
Malang	0.029411765	0
Pasuruan	0.117647059	0
Gresik	0.029411765	0
Surabaya	0.147058824	0
Lamongan	0.029411765	0
	1	1
P(Piutang)	Tidak Bermasalah	Sedikit Bermasalah
Ada	0	1
Tidak Ada	1	0
	1	1
P(Deadline Pembayaran)	Tidak Bermasalah	Sedikit Bermasalah
Tepat	1	0
Sedang	0	1
Tidak tepat	0	0

	1	1
P(Volume Transaksi)	Tidak Bermasalah	Sedikit Bermasalah
Sedikit	0.382352941	0
Sedang	0.117647059	0
Lumayan Banyak	0.176470588	0.5
Banyak	0.323529412	0.5
	1	1

Table 4. *Probabilitas*

Setelah mendapatkan nilai probabilitas, kemudian menerapkan Algoritma Naïve Bayes pada Data Testing. Berikut adalah rumus nya:

Mengkalikan probabilitas tiap variable dari setiap probabilitas label untuk menentukan nilai probabilitas terbesar sebagai prediksi untuk dibandingkan.

Contoh perhitungan seperti berikut :

Jika diketahui Data Testing seperti berikut : Pembayaran = Lunas, Potongan = Tidak, Daerah = Prambon, Piutang = Tidak Ada, Deadline Pembayaran = Tepat, Voume Transaksi = Sedikit, Status = Tidak Bermasalah. Penilaian probabilitas variable atau atribut sesuai data testing pada tabel perhitungan manual data transaksi seperti berikut :

$$P(\text{Pembayaran} = \text{Lunas} \mid \text{Status} = \text{Tidak Bermasalah}) = 1$$

$$P(\text{Pembayaran} = \text{Lunas} \mid \text{Status} = \text{Sedikit Bermasalah}) = 0$$

$$P(\text{Pembayaran} = \text{Lunas} \mid \text{Status} = \text{Sangat Bermasalah}) = 0$$

$$P(\text{Potongan} = \text{Tidak} \mid \text{Status} = \text{Tidak Bermasalah}) = 0.882352941$$

$$P(\text{Potongan} = \text{Tidak} \mid \text{Status} = \text{Sedikit Bermasalah}) = 1$$

$$P(\text{Potongan} = \text{Tidak} \mid \text{Status} = \text{Sangat Bermasalah}) = 1$$

$$P(\text{Daerah} = \text{Prambon} \mid \text{Status} = \text{Tidak Bermasalah}) = 0.058823529$$

$$P(\text{Daerah} = \text{Prambon} \mid \text{Status} = \text{Sedikit Bermasalah}) = 0$$

$P(\text{Daerah} = \text{Perambon} \mid \text{Status} = \text{Sangat Bermasalah}) = 0$

$P(\text{Piutang} = \text{Tidak Ada} \mid \text{Status} = \text{Tidak Bermasalah}) = 1$

$P(\text{Piutang} = \text{Tidak Ada} \mid \text{Status} = \text{Sedikit Bermasalah}) = 0$

$P(\text{Piutang} = \text{Tidak Ada} \mid \text{Status} = \text{Sangat Bermasalah}) = 0$

$P(\text{Deadline Pembayaran} = \text{Tepat} \mid \text{Status} = \text{Tidak Bermasalah}) = 1$

$P(\text{Deadline Pembayaran} = \text{Tepat} \mid \text{Status} = \text{Sedikit Bermasalah}) = 0$

$P(\text{Deadline Pembayaran} = \text{Tepat} \mid \text{Status} = \text{Sangat Bermasalah}) = 0$

$P(\text{Voume Transaksi} = \text{Sedikit} \mid \text{Status} = \text{Tidak Bermasalah}) = 0.382352941$

$P(\text{Voume Transaksi} = \text{Sedikit} \mid \text{Status} = \text{Sedikit Bermasalah}) = 0$

$P(\text{Voume Transaksi} = \text{Sedikit} \mid \text{Status} = \text{Sangat Bermasalah}) = 0$

Lalu mengkalikan nilai probabilitas tiap variable berdasarkan masing-masing target yang nanti akan dikalikan dengan nilai probabilitas tiap label.

$P(F \mid \text{Status} = \text{Tidak Bermasalah}) = 1 \times 0.882352941 \times 0.058823529 \times 1 \times 1 \times 1 \times$

$0.382352941 \times 0.80952381$

$= 0.01606525$

$P(F \mid \text{Status} = \text{Sedikit Bermasalah}) = 0 \times 1 \times 0 \times 0 \times 0 \times 0 \times 0 \times 0.142857143$

$= 0$

$P(F \mid \text{Status} = \text{Sangat Bermasalah}) = 0 \times 1 \times 0 \times 0 \times 0 \times 0 \times 0 \times 0.047619048 = 0$

Hasil Data Testing :

Status	PREDICTION	CLASS PREDICTION	Tidak Bermasalah	Sedikit Bermasalah	Sangat Bermasalah
Tidak Bermasalah		Tidak Bermasalah	0.01606525	0	0
Sangat Bermasalah		Sangat Bermasalah	0	0	0.023809524
Tidak Bermasalah		Tidak Bermasalah	0.00098863	0	0
Sedikit Bermasalah		Sedikit Bermasalah	0	0	0
Tidak Bermasalah		Tidak Bermasalah	0.03213050	0	0
Tidak Bermasalah		Tidak Bermasalah	0.00910364	0	0
Tidak Bermasalah		Tidak Bermasalah	0.04757785	0	0
Tidak Bermasalah		Tidak Bermasalah	0.04757785	0	0
Tidak Bermasalah		Tidak Bermasalah	0.00910364	0	0
Sangat Bermasalah		Sangat Bermasalah	0	0	0.023809524
Tidak Bermasalah		Tidak Bermasalah	0.00803262	0	0
Tidak Bermasalah		Tidak Bermasalah	0.00370737	0	0
Tidak Bermasalah		Tidak Bermasalah	0.01853683	0	0
Tidak Bermasalah		Tidak Bermasalah	0.02039051	0	0
Tidak Bermasalah		Tidak Bermasalah	0.02409787	0	0

Tidak Bermasalah		Tidak Bermasalah	0.04016312	0	0
Tidak Bermasalah		Tidak Bermasalah	0.03213050	0	0
Tidak Bermasalah		Tidak Bermasalah	0.00741473	0	0

Table 5. Hasil Data Testing

Dari hasil perbandingan tiap kelas dalam menentukan karakter pelanggan dengan class status. Nilai yang dihasilkan sesuai dengan data Transaksi Pelanggan yang dicari seperti berikut :

		CLASS		
Confusion Tabel	Predicted	Tidak Bermasalah	Sedikit Bermasalah	Sangat Bermasalah
	Tidak Bermasalah	15	0	0
	Sedikit Bermasalah	0	1	0
	Sangat Bermasalah	0	0	2
		Hasil Nilai Kebenaran	15	1
		Prosentase	83%	6%

Table 6. Confusion Tabel

Dari hasil yang ada pada Tabel 6. Confusion Tabel tercatat ada 15 data transaksi pelanggan dari Data Testing yang Tidak Bermasalah dengan prosentase 83%, sebanyak 1 data transaksi pelanggan yang Sedikit Bermasalah dengan prosentase 6%, dan 2 data transaksi pelanggan yang Sangat Bermasalah dengan prosentase 11%. Data tersebut diambil dari pencocokan nilai Benar dan salah pada kolom 'Status' dan kolom ' Class Prediction pada Tabel 5. Hasil Data Testing.

		CLASS			Prosentase
ACCURACY	Predicted	Tidak Bermasalah	Sedikit Bermasalah	Sangat Bermasalah	
	Benar	15	1	2	100%
	Salah	0	0	0	0%

Table 7. Accuracy Naïve Bayes

Dalam Tabel 7. Accuracy Naïve Bayes menunjukkan hasil akurasi dengan metode naïve bayes pada kelas Tidak Bermasalah sebesar 83% dengan nilai Benar sebanyak 15 data dan nilai salah sebanyak 0 data, kelas Sedikit Bermasalah sebesar 6% dengan nilai Benar sebanyak 1 data dan nilai Salah sebanyak 0 data, dan kelas Sangat Bermasalah sebesar 11% dengan nilai Benar sebanyak 2 data dan nilai Salah sebanyak 0 data. Besar prosentase prediksi benar sebesar 100%, sedangkan prosentase Salah sebesar 0%.

Gambar 2. Total Data Testing untuk Pemrograman Python pada WEB

Gambar 2. Total Data Testing untuk Pemrograman python pada WEB merupakan tampilan dari accuracy yang dihitung melalui pemrograman python dengan jumlah data pada kategori 'Sangat Bermasalah' sebanyak 14 data, 'Sedikit Bermasalah' sebanyak 16 dan 'Tidak Bermasalah' sebanyak 150 dari 180 data transaksi yang menjadi data testing pada implementasi pada WEB menggunakan python.

Implementasi program dengan bahasa pemrograman pyton

A. Implementasi Upload Data Training

Gambar 3. Upload Data Training

Pada Gambar 3. Upload Data Training merupakan hasil setelah user akan mengupload data training dengan nama file training.csv. Dalam tampilan ini terdapat beberapa bagian rincian data seperti nama relasi, atribut dan jumlah data dari file training..csv. File tersebut akan dibaca oleh machine learning dan akan ditampilkan pada tabel yang ada. Pada Pengupload an Data Training menggunakan sebesar 70% dari keseluruhan data set yang berjumlah 600 data transaksi yang berjumlah 420 data sebagai data training.

B. Implementasi Upload Training

Gambar 4. Upload Data Testing

Pada Gambar 4. Upload Data Testing merupakan hasil setelah user mengupload dengan meng-klik tombol upload untuk mengupload data testing yang disimpan pada format file testing.csv. Dalam tampilan ini terdapat beberapa bagian rincian data seperti nama relasi, atribut dan jumlah data dari file testing..csv. Data tersebut akan ditampilkan pada tabel dibawah informasi. Pada Pengupload an Data TEsting menggunakan sebesar 70% dari keseluruhan data set yang berjumlah 600 data transaksi yang berjumlah 180 data sebagai data testing.

C. Implementasi Hasil Klasifikasi

Gambar 5. Classification

Gambar 5. Classification merupakan tampilan hasil dari klasifikasi yang telah diproses oleh program setelah user meng-klik tombol Submit. Gambar 5. Classification menampilkan relasi training, jumlah data (inatanse), Attributes data training, dan juga class yang ditentukan. Lalu ada tampilan dari data training dan data training yang sudah dirubah menjadi bentuk biner untuk memudahkan proses perhitungan menggunakan naïve bayes pada python. Lalu dibawahnya menunjukkan tabel prediction dimana prediction merupakan hasil prediksi dari data testing (data uji). Kemudian ada hasil accuracy yang didapatkan dalam perhitungan menggunakan metode naïve bayes.

Gambar 6. Accuracy Data Uji

Gambar 6. Accuracy Data Uji merupa hasil akhir yang diperlukan dalam penelitian untuk mengetahui tingkat akurasi dari data yang telah diproses. Untuk menghitung akurasi yaitu dengan cara membagi jumlah sentimen benar dengan jumlah data tes. Untuk mengukur performa klasifikasi seperti pada gambar 6. Accuracy Data Uji juga menghitung precision, recall dan f1-score. Perhitungan precision dengan cara membagi true positive dengan hasil penjumlahan dari true positive dan false positif. Untuk perhitungan recall dapat dilakukan dengan membagi true positif dengan hasil penjumlahan true positive dan false negative. Sedangkan f1-score didapatkan dari hasil dari perkalian precision dan recall dibagi dengan hasil penjumlahan precision dan recall yang kemudian dikalikan dua. Dari proses perhitungan naïve bayes yang ditunjukan pada Gambar 6. Accuracy Data Uji, akurasi yang di dapat dari pengujian data tes yaitu sebesar 0.9 dengan precision dari class sangat bermasalah sebesar 0.57, recall sebesar 0.86 dan f1-score sebesar 0.69 dengan support data sebesar 14. Dari class Sedikit Bermasalah memiliki precicion sebesar 0.00, recall sebesar 0.00 dan f1-score sebesar 0.00 dengan support data sebesar 16. Sedangkan class Tidak Bermasalah memiliki precision sebesar 0.95, recall sebesar 1.00 dan f1-score sebesar 0.97 dengan support data sebesar 150. Hasil tersebut memiliki 51% nilai benar dan 49% nilai salah.

Kesimpulan

Dari hasil penelitian yang telah dilakukan, dapat diambil kesimpulan sebagai berikut:

- Accuracy

Pada hasil pengujian didapat accuracy sebesar 0.9 dengan precision dari class sangat bermasalah sebesar 0.57, recall sebesar 0.86 dan f1-score sebesar 0.69 dengan support data sebesar 14. Dari class Sedikit Bermasalah memiliki precision sebesar 0.00, recall sebesar 0.00 dan f1-score sebesar 0.00 dengan support data sebesar 16. Sedangkan class Tidak Bermasalah memiliki precision sebesar 0.95, recall sebesar 1.00 dan f1-score sebesar 0.97 dengan support data sebesar 150.

- Hasil Accuracy dapat dijadikan acuan sebagai dasar untuk mengambil keputusan oleh pihak UD Delta.

References

1. Miranda, E. (2018). Data Mining Dengan Metode Klasifikasi Naïve Bayes Untuk Mengklasifikasikan Pelanggan. *Infotech: Journal of Technology Information*, 4(1), 6-12. <https://doi.org/10.37365/it.v4i1.7>
2. Siska, Y. (2019). Penerapan Data Mining Dengan Algoritma Naïve Bayes Pelanggan Terhadap Pelayanan Servis Mobil (Studi Kasus : Katamso Service). 14, 195-199.
3. Jiawei, Han, Micheline, K. (n.d.). *Konsep dan Teknik Data Mining Data Preprocessing*. Buku.
4. Topics, U. (2007). Principles of Data Mining. In *Principles of Data Mining*. <https://doi.org/10.1007/978-1-84628-766-4>
5. Bash, E. (2015). Python Programming for the Absolute Beginner. In *PhD Proposal (Vol. 1)*.
6. Riza, & Irsan. (2013). Pengenalan HTML, CSS, dan Javascript.
7. Ii, B. A. B. (2005). Data Mining Data mining. *Mining of Massive Datasets*, 2(January 2013), 5-20. [https://www.cambridge.org/core/product/identifier/CBO9781139s058452A007/type/book_part_preprocessing_millennial_2_\(n.d.\)](https://www.cambridge.org/core/product/identifier/CBO9781139s058452A007/type/book_part_preprocessing_millennial_2_(n.d.))
8. Mita, S., Yamazoe, Y., Kamataki, T., & Kato, R. (1981). Metabolic activation of a tryptophan pyrolysis product, 3-amino-1-methyl-5H-pyrido[4,3-b]indole(Trp-P-2) by isolated rat liver nuclei. In *Cancer Letters (Vol. 14, Issue 3)*. [https://doi.org/10.1016/0304-3835\(81\)90152-X](https://doi.org/10.1016/0304-3835(81)90152-X)
9. Egziabher, T. B. G., & Edwards, S. (2013). Pedoman Penulisan Abstrak Dan Artikel Young Public Health Award Kategori Hasil Penelitian Bidang Kesehatan Masyarakat Indonesian Public Health Perspective Series (Iphps) 2016. *Africa's Potential for the Ecological Intensification of Agriculture*, 53(9), 1689-1699.
10. Gaikwad, S. S., & Adkar, P. (2019). A Review Paper on Bootstrap Framework. *Ire*, 2(10), 349-351.
11. Raharjo, Budi. 2017. *Belajar Otodidak Flask (Framework Python untuk Pengembangan Aplikasi WEB)*. Bandung : Informatika Bandung
12. Raharjo, Budi. 2019. *Mudah Belajar Python untuk Aplikasi Desktop dan WEB*. Bandung : Informatika Bandung
13. Ratnawati, F. (2018). Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter. *INOVTEK Polbeng - Seri Informatika*, 3(1), 50. <https://doi.org/10.35314/isi.v3i1.335>
14. Fitriani, A. S. (2019). Penerapan Data Mining Menggunakan Metode Klasifikasi Naïve Bayes untuk Memprediksi Partisipasi Pemilihan Gubernur. *JTAM (Jurnal Teori Dan Aplikasi Matematika)*, 3(2), 98-104.
15. Yuniarti, W. D., Faiz, A. N., & Setiawan, B. (2020). Identifikasi Potensi Keberhasilan Studi Menggunakan Naïve Bayes Classifier. *Walisongo Journal of Information Technology*, 2(1), 1. <https://doi.org/10.21580/wjit.2020.2.1.5204>