

Implementation of The Naïve Bayes Algorithm on The Demography of Gresik District to Measure Election Participation

[Implementasi Algoritma Naïve Bayes pada Demografi Kabupaten Gresik untuk Mengukur Partisipasi Pemilu]

Hawwani Muhdat Azzahrah*¹⁾, Arif senja Fitriani ²⁾

1) Program Studi Teknik Informatika, Universitas Muhammadiyah Sidoarjo, Indonesia

2) Program Studi Teknik Informatika, Universitas Muhammadiyah Sidoarjo, Indonesia

*Email Penulis Korespondensi: asfjim@umsida.ac.id

Abstract. In 1995 Indonesia held elections for the first time. With the highest number of participation, namely 91.5%, in multiparty elections, there were 36 registered voters. In some countries, elections themselves are a way to influence decisions on how a country can be governed. However, elections themselves cannot be separated from the politics of bribery because there is a lot of competition in the political world for power. In the 2014 Presidential Election, based on KPU data, 30.42% were registered as Golput voters, namely citizens who did not exercise their right to vote during the election. The size of the election results greatly influences several factors. Like the problem that will be discussed this time, there are 3,738 data that will be processed using the Naïve Bayes method in Implementing the Demography of Gresik Regency with the highest accuracy of 62.84% which is later expected to measure election participation and become an example of improving electoral rights.

Keywords - Data Mining; Demographics; Election; Golput; Naïve Bayes.

Abstrak. Pada tahun 1995 Indonesia mengadakan pemilu untuk pertama kalinya. Dengan jumlah partisipasi tertinggi yaitu 91,5% pada pemilihan multipartai sebanyak 36 jumlah pemilih yang terdaftar. Pada sebagian negara pemilu sendiri merupakan cara yang dilakukan untuk mempengaruhi keputusan bagaimana suatu negara dapat diatur. Akan tetapi pemilu sendiri tidak lepas dari politik suap karena banyaknya persaingan di dunia Politik untuk memperebutkan kekuasaan. Pada Pemilu Presiden 2014 berdasarkan data KPU sebanyak 30,42% terdata sebagai pemilih Golput, yaitu warga yang tidak menggunakan hak pilihnya saat pemilu. Besar kecilnya angka pada perolehan pemilu sangat berpengaruh pada beberapa faktor. Seperti permasalahan yang akan dibahas kali ini yaitu sebanyak 3,738 data yang akan diolah dengan penerapan metode Naïve Bayes dalam Mengimplementasi Demografi Kabupaten Gresik dengan perolehan akurasi tertinggi sebanyak 62,84% yang nantinya diharapkan untuk mengukur partisipasi pemilu dan menjadi contoh perbaikan pada hak pemilu.

Kata Kunci - Data Mining; Demografi; Pemilu; Golput; Naïve Bayes.

I. PENDAHULUAN

Indonesia mengadakan pemilu pertama kali pada tahun 1995. Pemilu adalah proses pemungutan suara dari pilihan rakyat untuk memilih seseorang menjadi pemimpin suatu politik atau wakilnya dalam pemerintahan. Di dalam pemilu, suatu hal yang diperjuangkan partai politik tidak selalu searah dengan kepentingan perorangan secara langsung, meskipun hak itu bersangkutan dengan kepentingan umum yang luas. Dalam masalah semacam ini, seorang pemilih yang memiliki sifat acuh tak acuh dan tidak toleran seringkali menarik diri dari pencatutan politik langsung karena cenderung berfikir hal tersebut tidak berhubungan dengan kepentingannya [1].

Pemilihan multipartai dengan jumlah partisipasi tertinggi 91,5% dari 39 juta pemilih yang terdaftar. Dengan total 28 partai politik yang memperoleh kursi tetapi hanya 4 yang memperoleh sekitar 75% suara. Dalam demokrasi sendiri pemilu merupakan cara utama bagi sebagian besar negara untuk mempengaruhi keputusan tentang bagaimana suatu negara dapat diatur dan warga mendapatkan kesempatan mereka ketika pemerintah memutuskan untuk mengadakan pemilihan [2].

Indonesia telah dianggap sebagai negara bebas dan adil. Akan tetapi, bangsa ini tidak lepas dari politik suap yang mana kekuasaan atau posisi politik dapat dibeli dengan uang. Karena masih banyaknya rakyat Indonesia yang memiliki kesadaran tentang pentingnya hak suara yang mereka berikan pada pemilu. Di dalam pemilu, suatu hal yang diperjuangkan partai politik tidak selalu searah dengan kepentingan perorangan secara langsung, meskipun hak itu bersangkutan dengan kepentingan umum yang luas. Dalam masalah semacam ini, seorang pemilih yang memiliki sifat acuh tak acuh dan tidak toleran seringkali menarik diri dari pencatutan politik langsung karena cenderung berfikir hal tersebut tidak berhubungan dengan kepentingannya [3].

Karena besarnya angka hak pemilih yang tidak menggunakan hak pilihnya (Golput), ketika diadakannya pemilihan umum tahun 2014 menurut data KPU jumlah pemilih dalam partisipasi pemilu sebanyak 75,11% dan 24,89%

memilih Golput. Sedangkan pada Pemilu Presiden 2014 sebesar 69,58% pemilih yang terdaftar, sebanyak 30,42% pemilih memilih Golput. Walaupun saat pemilu 2019 jumlah Golput menurun yakni sebanyak 22,5% [4]. Tapi besar kecilnya angka tersebut juga dapat berpengaruh terhadap beberapa faktor. Seperti masalah yang akan di bahas kali ini, peneliti ingin menerapkan metode klasifikasi Naïve Bayes untuk dapat mengimplementasi demografi Kabupaten Gresik untuk mengukur partisipasi pemilu.

Penelitian sebelumnya yang dilakukan oleh Kamil Erwansyah, Dicky Nofriansyah Mukhlis R, Tahun 2016 yang berjudul “Penguji-an klasifikasi Naïve Bayes untuk mengetahui ketertarikan pelanggan pada kartu internet XL”, menjelaskan tentang klasifikasi model untuk dapat hasil pembeda konsep atau tipe data yang bertujuan untuk memprediksi peluang masa depan dan dapat memprediksi laku tidaknya kartu internet baru agar perusahaan dapat meningkatkan strategi pemasaran [5]. Dan penelitian yang dilakukan Listyaningrum, n.d. pada Tahun 2015 dengan judul “Analisis Karakter DPT dengan metode klasifikasi Naïve Bayes untuk memperhitungkan partisipan pemilu”, menjelaskan tentang Partisipasi pemilu dalam memprediksi analisis DPT karena jumlah partisi pemilu yang menurun setiap tahun [6]. Keduanya memiliki kelebihan yang sama yaitu menghitung prediksi dari minat perorangan, dari penelitian. Bedannya penelitian Dicky yang memprediksi peluang minat kartu XL dimasa depan, sedangkan Listyaningrum memprediksi karakter DPT yang menurun pada pemilu. Kemudian pembeda dari penelitian tersebut dengan penelitian yang akan dilakukan ini adalah data yang digunakan. Jika penelitian terdahulu menggunakan data karakteristik perorangan, maka penelitian kali ini lebih menganalisa data dari Kabupaten Gresik yang sudah tersedia di laman KPU dan BPS.

Berdasarkan permasalahan yang ada peneliti berharap penelitian ini dapat menjadi tumpuan pemerintah dalam mengatasi masalah kesadaran masyarakat tentang hak-hak pemilu dan memiliki harapan bahwa hak suara yang mereka pilih tidak akan sia-sia.

Pada pengujian kali ini bertuju pada implementasi Data Mining untuk mengetahui jumlah atau tingkat partisipasi Pemilu Kabupaten Gresik menggunakan Metode klasifikasi Naive Bayes, sehingga nantinya akan mendapatkan menjadi informasi pada pemerintah atau pihak yang bersangkutan yang dapat menjadikan panduan tentang kesadaran masyarakat dengan hak suara golput.

II. METODE

Sebelum melakukan penelitian, pengumpulan data akan dilakukan terlebih dahulu. Dalam pengujian untuk memecahkan masalah. Berikut adalah pelaksanaan dalam proses pengumpulan data untuk melakukan penelitian ini:

A. Metode Mengumpulkan Data

a) Studi Literatur

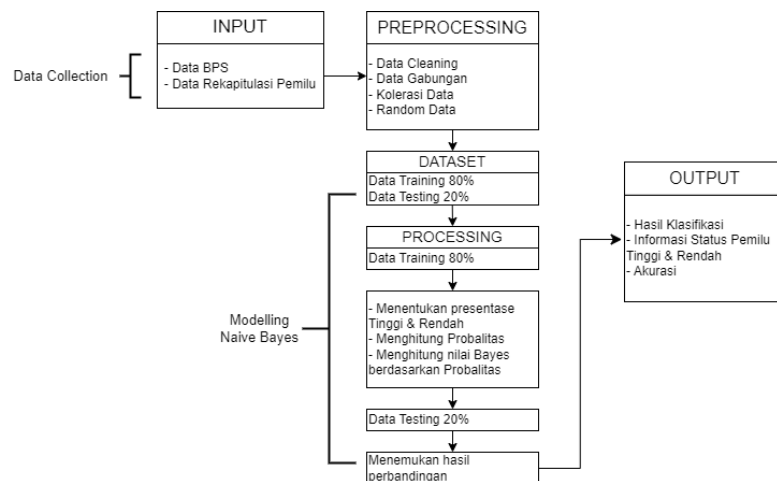
Pengumpulan data menggunakan metode studi kepustakaan ini melalui pencarian dari jurnal, buku, internet, dan literatur lainnya yang masih berkaitan dengan penelitian ini. Kemudian dari hasil literature yang didapatkan dalam pengumpulan data adalah adanya teori dasar yang diperoleh dapat digunakan sebagai penunjang dalam penelitian ini.

b) Observasi

Metode pengumpulan data dilakukan dengan pencarian dilaman web KPU dapat diakses pada <https://www.kpu.go.id/> dan BPS <https://www.bps.go.id/> . Kemudian dari data BPS yang tersedia data kemudian didownload, setelah itu untuk pengumpulan data KPU data yang diambil yaitu TPS, DPT, dan Hak Pilih. Informasi yang diperoleh dapat diterapkan dalam penelitian yang dilakukan.

B. Diagram Klasifikasi

Berikut merupakan proses diagram klasifikasi [7].



Gambar 1. Diagram Klasifikasi

C. Atribut Data

Pada penelitian ini terdapat sebanyak 18 atribut yang diambil dari data BPS dan 3 atribut dari kota Gresik. Atribut memerlukan adanya seleksi lagi, karena ada beberapa komponen yang ada pada tribute memiliki peran penting dalam studi ini, berikut penjelasannya pada tabel 1

Tabel 1. Atribut Data

No.	Atribut	Skala	Klarifikasi
1	Sistem Peringatan Dini Bencana Alam	Kategorikal	Ada, Tidak Ada
2	Sistem Peringatan Dini Khusus Tsunami	Kategorikal	Wilayah Berpotensi Tsunami, Bukan Wilayah Potensi Tsunami, Tidak Ada
3	Perlengkapan Keselamatan	Kategorikal	Ada, Tidak Ada
4	Rambu-rambu dan Jalur Evakuasi Bencana Alam	Kategorikal	Ada, Tidak Ada
5	Pembuatan, Perawatan, atau Normalisasi : Sungai, Kanal, Tanggul, Parit, Drainase, Waduk, Pantai, dll	Kategorikal	Ada, Tidak Ada
6	Minimarket/Swalayan	Kategorikal	Banyak, Sedikit
7	Toko/Warung Kelontong	Kategorikal	Banyak, Sedikit
8	Warung/Kedai Makan	Kategorikal	Banyak, Sedikit
9	Koperasi	Kategorikal	Banyak, Sedikit
10	Jumlah Menara Telepon Seluler (BTS)	Kategorikal	Banyak, Sedikit
11	Jumlah Operator Layanan Komunikasi Telepon Seluler yang Menjangkau di Desa/Kelurahan	Kategorikal	Banyak, Sedikit
12	Kondisi Sinyal Telepon Seluler di Sebagian Wilayah Desa/Kelurahan	Kategorikal	Sangat Kuat, Sinyal Kuat, Sinyal Lemah, Tidak Ada
13	Jenis Transportasi	Kategorikal	Darat, Air, Darat dan Air, Udara
14	Keberadaan Angkutan Umum	Kategorikal	Ada, Dengan trayek Tetap. Ada, Tanpa Trayek Tetap Tidak Ada Angkutan Umum
15	Jenis Permukaan Jalan	Kategorikal	Aspal/Beton, Diperkeras (Kerikil, Batu, dll)
16	Dapat Dilalui Kendaraan Roda 4 atau Lebih	Kategorikal	Sepanjang Tahun, Sepanjang Tahun Kecuali Kondisi Tertentu
17	Kantor Pos/ Pos Pembantu/ Rumah Pos	Kategorikal	Beroperasi, Jarang Beroperasi, Tidak Beroperasi, Tidak ada
18	Perusahaan/Agen jasa Ekspedisi Swasta	Kategorikal	Beroperasi, Jarang Beroperasi, Tidak ada

D. Input Data

Sebelum melakukan perhitungan data, data gabungan dari KPU dan BPS digabungkan sesuai dengan atribut yang telah diseleksi menggunakan excel. Langkah pengelompokan yang dilakukan berdasarkan atribut data yang di ambil dari data BPS sebanyak 18 atribut dan data KPU 3 atribut dengan total sebanyak 3,738 data yang nantinya akan dihitung. Kemudian untuk data golput dan hak pilih dilakukan Randomsize untuk nantinya menghasilkan 2 kelas atau label yaitu tinggi dan rendah sebagai acuan perhitungan. Dalam proses klasifikasi metode Naïve Bayes peneliti menggunakan 2 data untuk pengujian yaitu data training dan data testing dengan perbandingan 80:20 data.

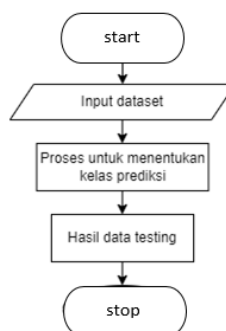
1) Data Training



Gambar 2. Flowchart Data Training

Menampilkan dataset setelah itu dilakukan proses pada data training berupa membaca data yang akan dihitung jumlah probalitasnya, kemudian dari proses tersebut akan menghasilkan tabel nilai bayes dengan probalitas, terakhir yaitu menghasilkan solusi penyelesaian. Dalam perhitungan probabilitas target atribut mempunyai karakter kategorikal dimana 17 atribut diambil dari dataset dan 1 atribut digunakan sebagai acuan untuk menentukan tinggi dan rendah.

2) Data Testing



Gambar 3. Flowchart Data Testing

Proses yang dilakukan pertama yaitu menginput data testing dari data probalitas sebelumnya, selanjutnya melakukan proses untuk menentukan kelas prediksinya, kemudian menghitung hasil perbandingan untuk memperoleh hasil pengujian. Pengujian data testing diaplikasikan untuk mengetahui seberapa dalamnya classifier berhasil dalam klasifikasi, lalu melihat hasil keakurasiannya dengan membaca per atribut dan tabel probalitas kemudian memasukan hasil prediksi.

E. Naïve Bayes

Untuk pengimplementasian data penulis menggunakan metode Naïve Bayes. Naive Bayes merupakan metode untuk mempelajari probalitas suatu objek yang memiliki ciri tertentu dari peninjauan sekumpulan data dengan menjumlahkan kombinasi dari dataset untuk menemukan hubungan dan meringkas data yang tidak di duga. Kemudahan dari proses ini adalah sekedar menggunakan data training dan data testing untuk menguji data yang dihasilkan [8]. Dengan penyusunan perangkat lunak/pengkodean yang berfokus menggunakan Python.

Berikut teorema metode Naïve Bayes :

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

Keterangan :

X : Data dengan klas yang masih belum ditemukan

H : Hipotesa data X adalah suatu klas yang tertentu

$P(H|X)$: Probabilitas hipotesa H untuk kondisi X (posteriori probabilitas)

$P(H)$: Probabilitas hipotesa H (prior probabilitas)

$P(X|H)$: Probabilitas X dalam kondisi hipotesis H

$P(X)$: Probalitas

III. HASIL DAN PEMBAHASAN

Pada bab ini terdapat pembahasan dari Implementasi Metode Naïve Bayes menggunakan pemograman Phyton. Dan menganalisa tahap proses dari pengujian, untuk mengetahui hasil dari partisipasi pemilu pada demografi desa Gresik.

F. Seleksi Data

Untuk menganalisis permasalahan dibutuhkan data yang telah dikumpulkan berbentuk tabel terlebih dahulu. Disini penulis telah mengumpulkan dat adari KPU dan BPS kota Gresik yang yang sudah dibahas sebelumnya. Kemudian data tersebut memlaui seleksi data yang menghasilkan 18 atribut gabungan. Dataset berupa file excel dengan 3,738 data yang telah di randomsize. Pada tabel 2 dan 3 merupakan contoh data sampel awal yang diambil

Tabel 2. Data Sampel Awal BPS

Kecamatan	Desa/Kelurahan	Jumlah Penduduk	Sistem Peringatan Dini Bencana Alam	Sistem Peringatan Dini Khusus Tsunami	Perlengkapan Keselamatan	Rambu-Rambu dan Jalur Evaluasi Bencana
Cerme	Banjarsari	7 732	2	2	2	2
Cerme	Betiting	5 694	2	2	2	2
Cerme	Cagakagung	2 273	2	2	2	2
Cerme	Cerme Kidul	6 023	2	2	2	2
Cerme	Cerme Lor	3 684	2	2	2	2

Tabel 3. Data Sampel Awal KPU

Kecamatan	Desa/Kelurahan	tps	dpt	hak pilih
ujung pangkah	sekapuk	1	262	217
		2	261	216
		3	261	222
		4	257	225
		5	258	215
		6	260	224
		7	258	221
		8	265	219
		9	268	220
		10	260	216
		11	259	202
		12	264	228
		13	260	206
		14	253	203

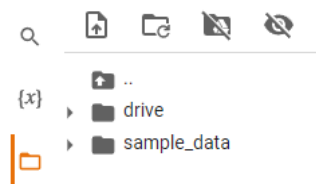
G. Input Dataset

Disini penulis menggunakan platform Google Colab. Maka dari itu dataset di unggah terlebih dahulu pada Drive, untuk mempermudah menghubungkan dataset dengan Google Colab nantinya.

```
#menghubungkan data pada drive
from google.colab import drive
drive.mount('/content/drive/')
```

Gambar 4. Menghubungkan Pada Drive

Setelah terhubung dengan Drive, langkah selanjutnya yaitu mencari data yang akan di gunakan pada file Drive, kemudian disalin pada coding. Seperti gambar 5.



Gambar 5. File Drive

```
import pandas as pd
# memanggil file excel pada drive
excel = pd.ExcelFile("/content/drive/MyDrive/DATA HAWWANI_raw.xlsx")

# sheet yang akan digunakan
df = pd.read_excel(excel, "HASIL RANDOMSIZE")

#menampilkan 5 data teratas
df.head()
```

Gambar 6. Source Code Menghubungkan Drive

Mounted at /content/drive/

	Sistem Peringatan Dini Bencana Alam	Sistem Peringatan Dini Khusus Tsunami	Perlengkapan Keselamatan	Rambu-rambu dan Jalur Evakuasi Bencana	Pembuatan, Perawatan, atau Normalisasi: Sungai, Kanal, Tanggul, Parit, Drainase, Waduk, Pantai, dll	Minimarket/Swalayan	Toko/Warung Kelontong	Warung/Kedai Makanan	Koperasi	Jumlah Menara Telepon Seluler (BTS)	Jumlah Operator Layanan Komunikasi Telepon Seluler Yang Menjangkau di Desa/Kelurahan	Kondisi Sinyal Telepon Seluler di sebagian Besar Wilayah Desa/Kelurahan1	Jenis Transportasi
0	1	0	0.2	1	1	0.272727	0.438202	0.375	0.014925	0.500	0.666667	0.0	0
1	1	0	0.2	1	1	0.000000	0.061798	0.035	0.029851	0.000	0.666667	0.5	0
2	1	0	0.2	1	1	0.454545	0.825843	0.435	0.059701	0.375	0.333333	0.0	0
3	1	0	0.0	1	1	0.000000	0.101124	0.160	0.029851	0.000	0.833333	0.5	0
4	1	0	0.2	1	1	0.000000	0.039326	0.060	0.014925	0.000	0.833333	0.0	0

Gambar 7. Menampilkan Dataset

H. Menentukan Label

Disini penulis menggunakan kolom "Label" yang berisi data Tinggi dan Rendah yang akan digunakan sebagai patokan prediksi data.

```
# menggunakan drop untuk meniadakan atau tidak mengikutkan kolom Label
data = df.drop('Label', axis=1)

# membuat target atau yang akan kita prediksi atau classnya dari kolom Label
target = df['Label']

print(f' data : {data.shape} \n target : {target.shape}')
```

Gambar 8. Source Code Menentukan Label

JUMLAH

1.00

Setelah mengetahui perhitungan naïve bayes dengan rumus, maka selanjutnya yaitu pengolahan data menggunakan pengkodean python. Peneliti menggunakan perulangan training untuk mengetahui hasil terbaik dari data yang telah di olah

```

# Build a Gaussian Naive Bayes Classifier
model = GaussianNB()

# Model training
model.fit(data_train, target_train)

# Make predictions on the test data
y_pred = model.predict(data_test)

# Calculate accuracy and F1 score
accuracy = accuracy_score(target_test, y_pred)
f1 = f1_score(target_test, y_pred)

# Append metrics to lists
accuracies.append(accuracy)
f1_scores.append(f1)

# Print the results for this iteration
print("Accuracy:", accuracy)
print("F1 Score:", f1)
print(classification_report(target_test, y_pred))
print("\n")

# Decrement the ratio for the next iteration
ratio -= 0.0125

# Print the average metrics across iterations
print("Average Accuracy:", sum(accuracies) / len(accuracies))
print("Average F1 Score:", sum(f1_scores) / len(f1_scores))

```

Gambar 11. Source Code Klasifikasi Naïve Bayes

Pada gambar 12 ini merupakan beberapa output dari klasifikasi.

```

ration: 0.1624999999999998
Iteration 20
data train : (3045, 18)
data test : (592, 18)
target train : (3045,)
target test : (592,)

Accuracy: 0.6283783783783784
F1 Score: 0.7465437788018433

```

	precision	recall	f1-score	support
0	0.60	0.20	0.30	236
1	0.63	0.91	0.75	356
accuracy			0.63	592
macro avg	0.62	0.56	0.53	592
weighted avg	0.62	0.63	0.57	592

Gambar 12. Output Klasifikasi Naïve Bayes

Dengan 20 output yang ditampilkan dari beberapa pengujian training maka kesimpulan dari iterasi yang di dapat untuk nilai presisi, recall, dan f1-support dalah sebagai berikut [10].

Tabel 5. Presisi, Recall, dan F1-Score

Iterasi	Presisi		Recall		F1-Score	
	0	1	0	1	0	1
1	0,53	0,60	0,19	0,88	0,27	0,71
2	0,53	0,60	0,19	0,88	0,28	0,71
3	0,53	0,60	0,18	0,88	0,27	0,71
4	0,53	0,60	0,17	0,89	0,26	0,72
5	0,53	0,60	0,16	0,90	0,25	0,72

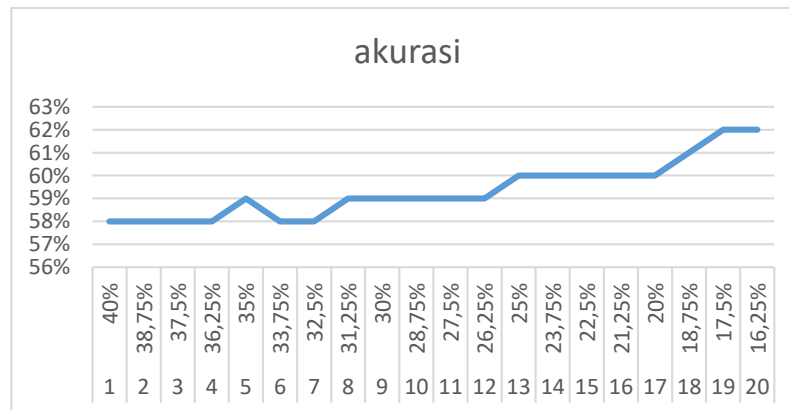
6	0,53	0,59	0,17	0,89	0,26	0,71
7	0,54	0,60	0,17	0,89	0,26	0,72
8	0,55	0,60	0,17	0,90	0,26	0,72
9	0,54	0,60	0,18	0,89	0,27	0,72
10	0,55	0,60	0,18	0,90	0,27	0,72
11	0,53	0,60	0,18	0,89	0,26	0,72
12	0,53	0,61	0,17	0,89	0,26	0,72
13	0,50	0,61	0,16	0,90	0,24	0,73
14	0,50	0,62	0,16	0,90	0,24	0,73
15	0,51	0,62	0,17	0,90	0,25	0,73
16	0,54	0,62	0,17	0,90	0,26	0,73
17	0,53	0,62	0,17	0,90	0,26	0,73
18	0,52	0,62	0,16	0,90	0,25	0,74
19	0,58	0,63	0,19	0,91	0,28	0,74
20	0,60	0,63	0,20	0,91	0,30	0,75

Selanjutnya yaitu tabel dari ratio ukuran berapa persen dari data testing keseluruhan data, dengan akurasinya.

Tabel 6. Ratio, dan Akurasi

Iterasi	Ratio	Akurasi
1	40%	58%
2	38,75%	58%
3	37,5%	58%
4	36,25%	58%
5	35%	59%
6	33,75%	58%
7	32,5%	58%
8	31,25%	59%
9	30%	59%
10	28,75%	59%
11	27,5%	59%
12	26,25%	59%
13	25%	60%
14	23,75%	60%
15	22,5%	60%
16	21,25%	60%
17	20%	60%
18	18,75%	61%
19	17,5%	62%
20	16,25%	62%

Pada gambar 13 menunjukkan hasil dari akurasi data train pada tabel sebelumnya.



Gambar 13. Grafik Akurasi

Langkah berikutnya yang dilakukan penulis yaitu membuat confusion matrix. Fungsi dari confusion matrix sendiri adalah dapat mengukur kinerja pada proses pelatihan atau training model. Dengan ciri khas yaitu menggunakan tabel 4 kolom yang memiliki perbedaan pada segi prediksi pada setiap kolomnya [11]. Dengan 4 nilai kombinasi pada gambar 14

		Actual Class	
		1	0
Predicted Class	1	True Positive	False Positive
	0	False Negative	True Negative

Gambar 14. Confusion Matrix

Confusion Matrix merupakan tabel yang digunakan untuk mengorganisir hasil prediksi model dengan membandingkannya dengan nilai actual dari data uji. Sangat berguna ketika ingin memahami seberapa baik model untuk mengklasifikasikan data ke dalam kategori yang berbeda, terutama dalam masalah klasifikasi. Dalam Confusion Matrix terdapat empat metrik utama yaitu True Positive (TP), False Positive (FP), True Negative (TN), dan False Negative (FN) [13].

Presisi

Presisi adalah ukuran berapa banyak dari contoh yang diprediksi secara positif yang benar-benar positif. Ini mengukur kemampuan model untuk menghindari false positif. Presisi yang tinggi menunjukkan bahwa ketika model memprediksi kelas positif, model tersebut sering kali benar. Dapat dihitung dengan persamaan (4)

$$\text{Presisi} = \frac{TP}{(TP+FP)} \times 100\% \quad (4)$$

Recall

Recall, juga dikenal sebagai sensitivitas atau tingkat positif sejati, mengukur kemampuan model untuk mengidentifikasi semua contoh positif dengan benar. Recall yang tinggi berarti model tersebut baik dalam menangkap sebagian besar contoh positif, meminimalkan negatif palsu. Dapat dihitung dengan persamaan (5)

$$\text{Presisi} = \frac{TP}{(TP+FP)} \times 100\% \quad (5)$$

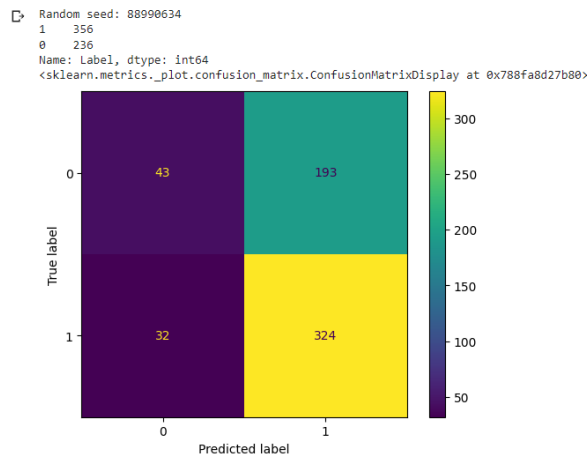
F1-Score

F1-Score adalah rata-rata harmonis dari presisi dan recall. Skor ini memberikan ukuran yang seimbang dari kinerja model, dengan mempertimbangkan positif palsu dan false negatif. F1-Score berguna ketika ingin menemukan keseimbangan antara presisi dan recall. Hal ini sangat berharga ketika ada ketidakseimbangan di antara kelas-kelas. Dapat dihitung dengan persamaan (6)

$$\text{F1-Score} = 2 \times \frac{(\text{Presisi} \times \text{Recall})}{(\text{Presisi} + \text{Recall})} \quad (6)$$

```
[19] from sklearn.metrics import confusion_matrix, ConfusionMatrixDisplay
# Assuming 'target_test' and 'y_pred' are defined earlier in your code
cm = confusion_matrix(target_test, y_pred)
# Create the ConfusionMatrixDisplay
disp = ConfusionMatrixDisplay(confusion_matrix=cm)
# Print random seed, value counts, and display the confusion matrix
print(f"Random seed: {seed}")
print(target_test.value_counts())
disp.plot()
```

Gambar 15. Source Code Confussion Matrix



Gambar 16. Hasil Confussion Matrix

Pada gambar (15) dan (16) merupakan source code untuk menampilkan tabel dari 4 gabungan nilai TP, FP, FN dan TN [14]. Sedangkan untuk mengetahui nilai dari presisi, recall dan f1 score [15] ada pada gambar (17)

```
from sklearn.metrics import precision_score, recall_score, f1_score
# Assuming 'target_test' and 'y_pred' are defined earlier in code
# Calculate Precision
precision = precision_score(target_test, y_pred)
# Calculate Recall
recall = recall_score(target_test, y_pred)
# Calculate F1 Score
f1 = f1_score(target_test, y_pred)
# Print the results
print("Precision:", precision)
print("Recall:", recall)
print("F1 Score:", f1)
```

Gambar 17. Source Code Presisi, Recall, F1 Score

```
Precision: 0.6266924564796905
Recall: 0.9101123595505618
F1 Score: 0.7422680412371134
```

Gambar 18. Output Presisi, Recall, F1 Score

Diambil dari jumlah akurasi terbesar diketahui yaitu sebanyak 0,62 maka terdapat nilai recall sebanyak 0,91 dan nilai F1 score yaitu 0,74. Model menunjukkan kemampuan yang cukup baik untuk mengidentifikasi positif yang benar (recall tinggi) dan menghindari membuat terlalu banyak prediksi positif yang salah (presisi yang layak). F1 Score menunjukkan bahwa model mencapai keseimbangan antara kedua aspek ini.

IV. Kesimpulan

Pada penelitian ini 3,738 data data berhasil dikumpulkan. Dengan 18 atribut yang berhasil diseleksi dari Data KPU dan BPS kota Gresik pada tahun 2021. Pada Iterasi 20 Ratio terbaik untuk perbandingan data train dan data test sebanyak 83,75% : 16,25%. Dan hasil akurasi yang didapatkan sebanyak 62,84%. Nilai Recall sebanyak 91%. Dan F1- Score sebanyak 74%. Beberapa faktor yang mempengaruhi akurasi adalah data cacat yang dapat menimbulkan gangguan, ketidak akuratan, sehingga melemahkan kemampuan model.

V. Saran

Ada beberapa faktor mengapa nilai akurasi kurang baik. Meningkatkan akurasi pada pengklasifikasian Naïve Bayes melibatkan kombinasi pemrosesan data [12], langkah yang dapat diambil untuk mengatasi masalah tersebut adalah : 1). pastikan pemrosesan data awal bersih, mengatasi nilai yang hilang dan melakukan normalisasi. 2). Sesuaikan hyperparameter klasifikasi, secara sistematis kumpulkan lebih banyak data. 3). melakukan uji coba dengan algoritma lain yang hasilnya nanti akan dilakukan perbandingan untuk memperbaiki kualitas data agar nilai keakuratan baik.

REFERENSI

- [1] J Bhakti, I.N., Yanuarti, S. and Nurhasim, M. (2009) Military politics, ethnicity and conflict in Indonesia. Oxford: University of Oxford. Centre for research on inequality, human security and ethnicity (CRISE). <https://assets.publishing.service.gov.uk/media/57a08b7bed915d3cfd000d4a/wp62.pdf>
- [2] T., K.C.S. and T., K.C.S. (2008) Hukum Tata Negara republik Indonesia: Pengertian Hukum Tata Negara Dan Perkembangan Pemerintahan Indonesia Sejak proklamasi kemerdekaan 1945 Hingga Kini. Jakarta: Rineka Cipta. <https://onesearch.id/Record/IOS3774.JAKPU00000000101372>.
- [3] N.H. Sardini "Pemilu di Indonesia," dalam Restorasi Penyelenggaraan Pemilu di Indonesia, edisi pertama, Yogyakarta, 2011, 1-7.
- [4] Djuyandi Y. (2014)Efektifitas Sosialisasi Politik Pemilihan Umum Legislatif Tahun 2014 Oleh Komisi Pemilihan Umum. BJP. Tersedia <https://doi.org/10.21512/humaniora.v5i2.3263>
- [5] Nofriansyah D, et al (2016) Penerapan Data Mining dengan Algoritma Naive Bayes Clasfier untuk Mengetahui Minat Beli Pelanggan terhadap Kartu Internet XL (Studi kasus di CV. Sumbee Utama , Telekomunikasi). SAINTIKOM. [online] 15(2), 81-92.
- [6] Listyaningrum, S. (2015.). Penerapan Data Mining Untuk Analisis Karakteristik DPT Non-Participate sebagai Prediksi Partisipan Pemilu dengan Menggunakan Metode Naive Bayes Classifier. CORE. [Online]
- [7] Fitriani, A.S. (2019) Penerapan Data Mining Menggunakan Metode Klasifikasi Naïve Bayes untuk Memprediksi Partisipasi Pemilihan Gubernur. JTAM. [Online]. 3(2), 98-104. Tersedia <https://doi.org/10.31764/jtam.v3i2.995>
- [8] Bayulianto S, et al. (2023) Prediksi Tingkat Kemenangan Mobile Legends Profesional League Indonesia Season o dengan Menggunakan Algoritma Naïve Bayes. JIPI. Tersedia <https://doi.org/10.29100/jipi.v8i2.3562>
- [9] Chowiyah, A. Penerapan Data Mining Menggunakan Metode Klasifikasi Naïve Bayes Untuk Memprediksi Partisipasi Pemilihan Gubernur dan Wakil Gubernur di Desa Jemirahan, Kecamatan Jabon, Univ. Umsida, Sidoarjo, Indonesia, 2019.
- [10] Hikmahtiyar S. Klasifikasi Gerakan Tangan Pada Protesis Untuk Pasien Amputasi dengan Artificial Intelligence. Univ. Airlangga, 2023.
- [11]Negoro N,K, et al (2023) Klasifikasi Deteksi Penggunaan Masker Menggunakan Metode Convolutional Neutral Network. JIPI. Tersedia <https://doi.org/10.29100/jipi.v8i2.3748>
- [12]Amelia, M. winny, Lumenta, A. S. ., & Jacobus, A. (2017). Prediksi Masa Studi Mahasiswa dengan Menggunakan Algoritma Naïve Bayes. Jurnal Teknik Informatika, [Online] 11(1), 1-10. Tersedia <https://doi.org/10.35793/jti.11.1.2017.17652>
- [13]Pratiwi B,N. et al. (2020). Pengukuran Kinerja Sistem Kualitas Udara Dengan Teknologi WSN Menggunakan Confusion Matrix. JTU. [Online] Tersedia <https://doi.org/10.26877/jiu.v6i2.6552>
- [14]Scikit-Learn "sklearn.metrics.confusion_matrix," 2023 [Online]. Aavailable https://scikit-learn.org/stable/modules/generated/sklearn.metrics.confusion_matrix.html [Diakses 20 Juni 2023]
- [15]Borwnlee J. Machine Learning Mastery "How to Calculate Precision, Recall, F1 and More for Deep Learning Models" 19 08 2022 [Online]. Aavailable <https://machinelearningmastery.com/how-to-calculate-precision-recall-f1-and-more-for-deep-learning-models/>. [Diakses 25 September 2023]

Conflict of Interest Statement:

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.