



# Implementasi Data Mining Untuk Memprediksi Calon Legislatif Terpilih Daerah Pemilihan Provinsi Banten Dengan *Algoritma Naive Bayes Classification*

Geugeut Ira Astria<sup>1</sup>

<sup>1</sup>Ilmu Komputer, Teknik Informatika, Universitas Pamulang, Tangerang Selatan, Indonesia

Email: [g.astria99@gmail.com](mailto:g.astria99@gmail.com)

**Abstrak** – Indonesia merupakan salah satu negara yang menganut sistem pemerintahan Presidensial di mana dalam memilih para calon pemimpin dan wakil rakyat di Indonesia biasanya diselenggarakan Pemilu (Pemilihan Umum). Tujuan dari penelitian ini adalah untuk memanfaatkan data mining dan metode Naive Bayes dalam memprediksi calon anggota legislatif terpilih di daerah pemilihan provinsi Banten pada tahun 2019. Data yang digunakan dalam penelitian ini diperoleh dari website KPU mengenai data calon anggota legislatif (DPR RI dan DPRD Provinsi) daerah pemilihan provinsi Banten berdasarkan tahun 2019. Hasil penelitian menunjukkan tingkat akurasi mencapai 95%, hasil yang sangat baik yang dapat memberikan informasi strategis bagi calon anggota legislatif dan partai politik dalam kampanye pemilu. Dengan penelitian data mining dengan Naive Bayes diharapkan dapat meningkatkan efektivitas kampanye pemilu dan memberikan wawasan yang bermanfaat bagi pemilihan umum di masa depan.

**Kata Kunci:** Data Mining, Naive Bayes, Pemilihan Umum, Anggota Legislatif, Prediksi, Akurasi.

*Abstract – Indonesia adheres to a Presidential system of government, where the selection of potential leaders and representatives typically involves the holding of General Elections (Pemilu). The purpose of this research is to utilize data mining and the Naive Bayes method to predict the elected legislative candidates in the provincial electoral districts of Banten in the year 2019. The data for this study was obtained from the KPU website, specifically focusing on the data of legislative candidates (DPR RI and DPRD Provincial) in the provincial electoral districts of Banten for the year 2019. The research results demonstrate a remarkable accuracy rate of 95%, providing valuable strategic insights for legislative candidates and political parties during election campaigns. Through data mining research employing the Naive Bayes method, it is anticipated that campaign effectiveness can be enhanced, offering valuable insights for future elections.*

**Keywords:** Data Mining, Naive Bayes, General Elections, Legislative Candidates, Prediction, Accuracy.

## 1. PENDAHULUAN

Indonesia merupakan salah satu negara yang menganut sistem demokrasi dengan sistem pemerintahan Presidensial di mana adanya pemisahan kekuasaan yaitu Eksekutif, Legislatif dan Yudikatif. Dalam memilih para calon pemimpin dan wakil rakyat di Indonesia biasanya diselenggarakan Pemilu (Pemilihan Umum).

Sesuai dengan ketentuan yang ada pada Pasal 22E ayat (6) Undang-Undang Dasar Negara Republik Indonesia Tahun 1945, pemilihan umum untuk memilih anggota Dewan Perwakilan Rakyat, Dewan Perwakilan Daerah, dan Dewan Perwakilan Rakyat Daerah diselenggarakan berlandaskan azas langsung, umum, bebas, rahasia, jujur, dan adil setiap lima tahun sekali. Pemilihan umum dimaksud diselenggarakan dengan menjamin prinsip keterwakilan, yang artinya setiap orang Warga Negara Indonesia terjamin memiliki wakil yang duduk di lembaga perwakilan yang akan menyuarakan aspirasi rakyat di setiap tingkatan pemerintahan, dari pusat hingga, ke daerah.

Pada penelitian ini data yang digunakan berupa data pemilu legislatif anggota Dewan Perwakilan Rakyat (DPR). Dalam teori trias politika yang dikemukakan Montesquieu, Lembaga Legislatif merupakan wakil rakyat yang diberikan kekuasaan untuk membuat undang-undang dan menetapkannya. Hal serupa juga dikemukakan oleh Miriam Budiardjo bahwa lembaga legislatif atau legislature mencerminkan salah satu tugas badan tersebut, yaitu legislate atau membuat undang-undang. (M.Hum, 2020)



Dalam pemilihan umum tentunya dibutuhkan strategi untuk memperoleh suara rakyat sebanyak mungkin agar dapat memenangkan pemilu tersebut. Ada banyak cara yang dilakukan baik oleh calon anggota legislatif maupun partai pengusung, mulai dari kampanye hingga melakukan promosi untuk mendapat suara pada pemilihan umum, seperti kampanye iklan di televisi yang dapat mempengaruhi pilihan individu.(Siregar dkk., 2019). Tentunya dalam melakukan kegiatan kampanye dibutuhkan data untuk diolah sebagai informasi agar kampanye yang dilakukan dapat tepat sasaran dan menuai hasil yang baik. Salah satu cara pengolahan data yang dapat dilakukan pada era kemajuan teknologi saat ini ialah dengan data mining.

Data mining sendiri merupakan serangkaian proses menggali atau menambang suatu kumpulan data untuk menghasilkan pengetahuan yang selama ini tidak diketahui secara manual data mining adalah proses untuk mendapatkan informasi yang berguna dari basis data yang besar dan perlu diekstraksi agar menjadi informasi baru dan dapat membantu dalam pengambilan keputusan.(Suntoro, 2019).

Hal ini sejalan dengan penelitian yang akan dilakukan mengenai prediksi untuk calon anggota legislatif dan untuk mendapatkan atau menggali data dengan tingkat akurasi yang lebih baik, dalam implementasi data mining pada penelitian ini metode yang digunakan adalah metode Naive Bayes.

Naive Bayes merupakan Teknik prediksi berbasis probabilistic sederhana yang berdasar pada penerapan teorama Bayes (atau aturan Bayes) dengan asumsi independensi (ketidaktergantungan) yang kuat (naif). Dengan kata lain, dalam Naive Bayes, model yang digunakan adalah “model fitur independen”. Dalam Bayes (terutama Naive Bayes), maksud independensi yang kuat pada fitur adalah bahwa sebuah fitur pada sebuah data tidak berkaitan dengan nada atau tidaknya fitur lain dalam data yang sama.(Sanubari dkk., 2020).

Dalam penelitian ini salah satu teknik dari data mining yang digunakan yaitu Klasifikasi (Classification) yang merupakan suatu teknik untuk menilai objek data serta mengelompokkan objek berdasarkan atribut - atribut atau ciri objek ke dalam salah satu kategori yang telah didefinisikan.(Ramadhan & Andarsyah, 2022)

Pada penelitian ini data yang akan digunakan merupakan data yang tersedia pada website KPU yang berupa data calon anggota legislatif (DPR RI dan DPRD Provinsi) daerah pemilihan provinsi Banten berdasarkan tahun yang paling terkini yaitu tahun 2019 yang akan dibagi menjadi data training dan juga data testing.

## **2. METODOLOGI PENELITIAN**

### **2.1 Metode Penelitian**

Metode yang digunakan dalam penelitian ini adalah metode penelitian sekunder atau secondary research ialah jenis penelitian yang melakukan pengumpulan data dari sumber yang sudah ada, seperti artikel, laporan, jurnal dan juga basis data elektronik. Metode penelitian sekunder digunakan untuk mendapatkan informasi yang telah diterbitkan sebelumnya dan dapat digunakan untuk mendukung penelitian yang sedang dilaksanakan atau untuk melengkapi penelitian yang sudah ada.(Ibrahim dkk., 2023)

### **2.2. Metode Pengumpulan data**

Metode pengumpulan data adalah teknik ataupun cara yang digunakan penulis untuk mengumpulkan data yang nantinya data tersebut akan digunakan oleh penulis untuk memperoleh bahan, keterangan, dan informasi terkait dengan penelitian ini. Imron, I. (2019). Pada tahap pengumpulan data metode yang digunakan pada penelitian adalah :

#### **1. Studi Pustaka**

Studi pustaka dilakukan dengan cara membaca, mengutip, dan membuat catatan yang bersumber pada bahan-bahan pustaka yang mendukung dan berkaitan dengan penelitian yang

menggunakan metode Naive Bayes. Lalu selanjutnya mempelajari serta memahami jurnal dan buku-buku yang dijadikan referensi, yang berhubungan dengan penelitian yang akan dibahas. Maka hasil dari studi pustaka ini dicantumkan dalam landasan teori. Hal ini dimaksudkan agar penelitian memiliki landasan teori yang kuat untuk menarik kesimpulan.

## 2. Data Sekunder

Data Sekunder yaitu penggunaan data yang sudah ada jadi peneliti tidak langsung menerima dari sumber data. Secara singkat dapat dikatakan bahwa data sekunder adalah data yang telah dikumpulkan oleh pihak lain. Peneliti dapat mencari data sekunder ini melalui sumber data sekunder.(Ph.D dkk., 2023). Dalam penelitian ini data yang digunakan merupakan data sekunder. Data sekunder merupakan sumber data yang didapatkan peneliti dengan media perantara atau tidak secara langsung. Data sekunder yang diperoleh dari website milik KPU merupakan data pemilihan umum calon anggota legislatif (DPR RI dan DPRD Provinsi) daerah pemilihan provinsi Banten tahun 2019. Data Pemilu sendiri merupakan data publik karena data tersebut dipublikasikan oleh KPU dan data tersebut dapat diakses oleh siapa saja.

## 3. ANALISA DAN PEMBAHASAN

### 3.1 Analisa

#### 3.1.1 Persiapan Data

Pada tahap ini diawali dengan melakukan pengambilan data sampel atau contoh data dari data calon anggota legislatif tahun 2019 yang nantinya akan dijadikan sebagai data training dan data testing. Data yang digunakan sudah dilakukan proses pembersihan data dan transformasi data dalam bentuk kategori. Pengujian menggunakan data calon anggota legislatif tahun 2019. Data calon anggota legislatif sebanyak 1399 record diambil 47 record untuk digunakan sebagai data training dan 13 sebagai data testing yang total keseluruhannya adalah 50 data untuk sample uji coba metode naive bayes. Berdasarkan hasil pengolahan data dan jumlah data yang digunakan tersebut dapat dibagi menjadi dua kelas kategori yaitu terpilih sebanyak 20 anggota dan tidak terpilih sebanyak 30 anggota. Pada proses pengujian, data dibagi menjadi 2 bagian yaitu data training dan data testing dengan menggunakan metode algoritma Naive Bayes. Data training digunakan untuk membentuk tabel probabilitas dan data testing digunakan untuk menguji probabilitas yang telah terbentuk. Data tersebut dapat dilihat pada tabel dibawah ini:

Tabel 1. Data Training.

NO	NAMA	PARTAI POLITIK	JENIS KELAMIN	SUARA SAH CALEG	SUARA SAH PARTAI	DAERAH PEMILIHAN	KEANGGOTAAN	HASIL
1	WAHYU WIDIYA SURYANI NING WALUYO	PKB	P	1063	42514	BANTEN I	DPR RI	TIDAK TERPILIH
2	HJ. TATI NURCAHYANA	Gerindra	P	26687	130413	BANTEN 02	DPRD PROVINSI	TERPILIH
3	INDAH RUSMIATI	PDIP	P	12837	123685	BANTEN 03	DPRD PROVINSI	TERPILIH
4	Drs. H. MUJAKKIR ZUHRI	Golkar	L	22267	106725	BANTEN 04	DPRD PROVINSI	TERPILIH
5	RIA MAHDIA FITRI S.SOS	Nasdem	P	8080	41068	BANTEN 07	DPRD PROVINSI	TERPILIH
6	RESTI HERAWATI	Garuda	P	698	3760	BANTEN 09	DPRD PROVINSI	TIDAK TERPILIH
7	dr. HJ. SHINTA WISHNU	PKS	P	13314	27783	BANTEN 10	DPRD PROVINSI	TERPILIH
8	RAE PUTRI S.E	Perindo	P	752	9095	BANTEN 06	DPRD PROVINSI	TIDAK TERPILIH
9	LAURA NATALIA SINAGA SH	PSI	P	5789	23485	BANTEN 05	DPRD PROVINSI	TIDAK TERPILIH
10	HJ. ELLA SILVIA	PAN	P	14264	34838	BANTEN 06	DPRD PROVINSI	TERPILIH
11	Dr. INNE IRAWATY, SpKJ	PBB	P	1064	5700	BANTEN II	DPR RI	TIDAK TERPILIH
12	ALFA ANGELICA E SEREH	PKPI	P	272	1193	BANTEN II	DPR RI	TIDAK TERPILIH
13	MARTINA, S.I.KOM., M.Si	Gerindra	P	28539	209098	BANTEN III	DPR RI	TERPILIH

14	ALI ZAMRONI,	Gerindra	L	56792	80862	BANTEN I	DPR RI	TERPILIH
15	H. UMAR BIN BARMAWI ST MM	PKB	L	10242	43638	BANTEN 02	DPRD PROVINSI	TERPILIH
16	Ir. SURYADI HENDARMAN,	PKB	L	19009	29041	BANTEN II	DPR RI	TIDAK TERPILIH
17	MOH. BAHRI S.Pd.I, S.H.	Gerindra	L	19602	116177	BANTEN 03	DPRD PROVINSI	TERPILIH
18	ST. ANANTA WAHANA, S.H.	PDIP	L	26662	195074	BANTEN III	DPR RI	TERPILIH
19	UBALINUS SINULINGGA S.IP	PDIP	L	7160	135174	BANTEN 07	DPRD PROVINSI	TIDAK TERPILIH
20	DR. H. FURTASAN ALI YUSUF SE.	Nasdem	L	25729	35597	BANTEN 01	DPRD PROVINSI	TERPILIH
21	YARFARLINANDA	Garuda	L	358	4197	BANTEN 04	DPRD PROVINSI	TIDAK TERPILIH
22	H. ANHAR, SE	Berkarya	L	4754	14844	BANTEN I	DPR RI	TIDAK TERPILIH
23	DR. H. JAZULI JUWAINI, MA	PKS	L	68538	37534	BANTEN II	DPR RI	TERPILIH
24	H. IIP MIFTAHUL CHOIRY, S.PdI	PPP	L	49993	27494	BANTEN I	DPR RI	TERPILIH

### 3.1.2 Analisis Metode Naive Bayes

#### a. Menghitung Jumlah dan Probabilitas

Probabilitas Kelas :

**Tabel 2.** Probabilitas Kelas.

CLASS		PROBABILITAS	
TERPILIH	TIDAK TERPILIH	TERPILIH	TIDAK TERPILIH
17/37	20/37	0,4595	0,5405

Probabilitas Atribut :

**Tabel 3.** Probabilitas Partai Politik.

PARTAI POLITIK	TERPILIH	TIDAK TERPILIH	PROBABILITAS	
			TERPILIH	TIDAK TERPILIH
PKB	1	2	0,0588235	0,1000000
Gerindra	4	1	0,2352941	0,0500000
PDIP	2	1	0,1176471	0,0500000
Golkar	1	0	0,0588235	0,0000000
Nasdem	2	1	0,1176471	0,0500000
Garuda	0	3	0,0000000	0,1500000
Berkarya	0	2	0,0000000	0,1000000
PKS	2	0	0,1176471	0,0000000
Perindo	0	3	0,0000000	0,1500000
PPP	1	1	0,0588235	0,0500000
PSI	0	1	0,0000000	0,0500000
PAN	2	1	0,1176471	0,0500000
Hanura	0	1	0,0000000	0,0500000
Demokrat	2	0	0,1176471	0,0000000
PBB	0	1	0,0000000	0,0500000
PKPI	0	2	0,0000000	0,1000000
Jumlah	17	20	1	1

#### b. Menghitung Nilai Mean

Dalam penelitian ini digunakan dua kelas yaitu Terpilih dan Bad Cust. Sebelum mencari nilai mean dan standar deviasi, kita harus mengelompokkan terlebih dahulu jumlah kelas Terpilih dan



Bad Cust berdasarkan data training yang digunakan. Untuk menghitung nilai mean di gunakan persamaan (5). Perhitungannya dapat dilihat sebagai berikut:

$$\text{Terpilih} = 6$$

$$\text{Tidak Terpilih} = 24$$

Menghitung Nilai Mean

$$\mu \text{ (Terpilih)} = \frac{(26687+12837+22267+8080+13314+14264+28539+56792+10242+19602+26662+25729+68538+49993+62509+60064+12352)}{17}$$

$$\mu \text{ (Terpilih)} = (518471)/17$$

$$\mu \text{ (Terpilih)} = 30498,294$$

$$\mu \text{ (Tidak Terpilih)} = \frac{(1063+698+752+5789+1064+272+19009+7160+358+4754+691+101+17125+14886+1039+191+4044+1965+521+193)}{20}$$

$$\mu \text{ (Tidak Terpilih)} = (81675)/20$$

$$\mu \text{ (Tidak Terpilih)} = 4083,75$$

Data diatas diambil dari data pada variabel suara sah caleg dengan hasil di kelas Terpilih, sehingga didapatkan nilai  $\mu$  nya. Untuk perhitungan nilai  $\mu$  dengan hasil Tidak Terpilih pada variabel suara sah caleg langkah-langkah nya sama. Sedangkan untuk menghitung nilai  $\mu$  pada variabel lainnya langkah-langkah yang digunakan juga sama seperti menghitung nilai mean pada variabel suara sah caleg.

### c. Menghitung Standar Deviasi

Setelah mendapatkan nilai  $\mu$  pada masing-masing variabel, maka selanjutnya kita mencari nilai standar deviasi, Untuk menghitung nilai standar deviasi digunakan persamaan (2). Berikut adalah proses menghitung nilai standar deviasi.

$$\sigma \text{ (Terpilih)} = \sqrt{\frac{(26687 - 30498,294)^2 + (12837 - 30498,294)^2 + (22267 - 30498,294)^2 + (8080 - 30498,294)^2 + (13314 - 30498,294)^2 + (14264 - 30498,294)^2 + (28539 - 30498,294)^2 + (56792 - 30498,294)^2 + (10242 - 30498,294)^2 + (19602 - 30498,294)^2 + (26662 - 30498,294)^2 + (25729 - 30498,294)^2 + (68538 - 30498,294)^2 + (49993 - 30498,294)^2 + (62509 - 30498,294)^2 + (60064 - 30498,294)^2 + (12352 - 30498,294)^2}{(17 - 1)}}$$

$$\sigma \text{ (Terpilih)} = \sqrt{6772508482/16}$$

$$\sigma \text{ (Terpilih)} = \sqrt{423281780,1}$$

$$\sigma \text{ (Terpilih)} = 20573,813$$

$$\sigma \text{ (Tidak Terpilih)} = \sqrt{\frac{(1063 - 4083,75)^2 + (698 - 4083,75)^2 + (752 - 4083,75)^2 + (5789 - 4083,75)^2 + (1064 - 4083,75)^2 + (272 - 4083,75)^2 + (19009 - 4083,75)^2 + (7160 - 4083,75)^2 + (358 - 4083,75)^2 + (4754 - 4083,75)^2 + (691 - 4083,75)^2 + (101 - 4083,75)^2 + (17125 - 4083,75)^2 + (14886 - 4083,75)^2 + (1039 - 4083,75)^2 + (191 - 4083,75)^2 + (4044 - 4083,75)^2 + (1965 - 4083,75)^2 + (521 - 4083,75)^2 + (193 - 4083,75)^2}{(20 - 1)}}$$

$$\sigma \text{ (Tidak Terpilih)} = \sqrt{675683513,8/19}$$

$$\sigma \text{ (Tidak Terpilih)} = \sqrt{35562290,2}$$

$$\sigma \text{ (Tidak Terpilih)} = 5963,413$$

Setelah mencari nilai probabilitas tiap fitur, maka didapat tabel untuk nilai, mean, standar deviasi dan probabilitas untuk masing-masing variabel, bisa dilihat pada tabel berikut:

**Tabel 4.** Probabilitas Suara Sah Caleg.

NILAI	TERPILIH	TIDAK TERPILIH
$\mu$ MEAN	30498,2941	4083,75
$\sigma$ DEVIASI <sup>2</sup>	423281780,0956	35562290,1974
$\sigma$ DEVIASI	20573,8130	5963,4126

**Tabel 5.** Probabilitas Suara Sah Partai.

NILAI	TERPILIH	TIDAK TERPILIH
$\mu$ MEAN	80738,1765	25912,95
$\sigma$ DEVIASI <sup>2</sup>	3388432394,1544	1103140193,9447
$\sigma$ DEVIASI	58210,2430	33213,5544

**d. Menghitung Nilai Distribusi Gaussian**

Langkah selanjutnya adalah menghitung nilai distribusi gaussian untuk data testing yang mempunyai data numerik/angka.

**Tabel 6.** Data Testing.

NO	NAMA	PARTAI POLITIK	JENIS KELAMIN	SUARA SAH CALEG	SUARA SAH PARTAI	DAERAH PEMILIHAN	KEANGGOTAAN	HASIL
1	AGUS EFENDI	PPP	L	10103	35308	BANTEN 01	DPRD PROVINSI	TERPILIH
2	H. MUHAMMAD RIZAL, SH, M.Si	PAN	L	18617	53192	BANTEN III	DPR RI	TERPILIH
3	GOUSTA FERIZA, S.H., M.H	Golkar	L	4527	42286	BANTEN I	DPR RI	TIDAK TERPILIH
4	KOKOK H. DIRGANTORO	PSI	L	7746	43706	BANTEN III	DPR RI	TIDAK TERPILIH
5	ARI GUNAWAN	PKPI	L	512	3170	BANTEN III	DPR RI	TIDAK TERPILIH
6	DESY YUSANDI S.E	Golkar	P	14988	35155	BANTEN 06	DPRD PROVINSI	TERPILIH
7	AKLINA NOVA RIKA S.IP	PSI	P	363	3295	BANTEN 08	DPRD PROVINSI	TIDAK TERPILIH
8	DWI UTAMI, SH	Hanura	P	831	15411	BANTEN III	DPR RI	TIDAK TERPILIH
9	DR. IR. SYLVIA WISNA DARWIS, SE., M.M.	PBB	P	2300	12049	BANTEN III	DPR RI	TIDAK TERPILIH
10	ZULHIDJAYANTO HUSIN	PKS	L	2356	62736	BANTEN 05	DPRD PROVINSI	TIDAK TERPILIH
11	HJ. MARYANA SE	Berkarya	P	1924	16435	BANTEN 08	DPRD PROVINSI	TIDAK TERPILIH
12	ASTRID OCTAVIN SARMARIA	Hanura	P	628	5456	BANTEN II	DPR RI	TIDAK TERPILIH
13	DRS. MUKLIS MM	PBB	L	1069	16448	BANTEN 03	DPRD PROVINSI	TIDAK TERPILIH

Untuk menghitung nilai distribusi gaussian, menggunakan persamaan, sebagai berikut adalah proses menghitung nilai distribusi Gaussian:

$$P(\text{Suara Sah Caleg} = 10103|\text{Terpilih}) = \frac{1}{\sqrt{2\pi} \cdot 20573,8130} * e^{-\frac{(10103-30498,2941)^2}{(2 \cdot 423281780,0956)}}$$

$$P(\text{Suara Sah Caleg} = 10103|\text{Terpilih}) = \frac{1}{129203,5454} * 2,71828183^{-0,491360651}$$

$$P(\text{Suara Sah Caleg} = 10103|\text{Terpilih}) = \frac{1}{129203,5454} * 0,61179339$$

$$P(\text{Suara Sah Caleg} = 10103|\text{Terpilih}) = 0,001702031$$

**e. Probabilitas akhir setiap kelas**

Menghitung nilai probabilitik dengan cara menghitung jumlah data yang sesuai dari kategori yang sama dibagi dengan jumlah data pada kategori tersebut, dapat dilihat menggunakan persamaan, sebagai berikut:



P(X|Terpilih) = P(Partai Politik | Terpilih)\*P(Jenis Kelamin | Terpilih)\*P(Suara Sah Caleg | Terpilih)\*P(Suara Sah Partai | Terpilih)\*P(Daerah Pemilihan | Terpilih)\* P(Keanggotaan | Terpilih)

P(X|Terpilih) = 0,0588235\*0,6470588\*0,0017020\*0,0012197\*0,0588235\*0,5882353

P(X|Terpilih) = 0,0000000027341

P(X|Tidak Terpilih) = P(Partai Politik | Tidak Terpilih)\*P(Jenis Kelamin | Tidak Terpilih)\*P(Suara Sah Caleg | Tidak Terpilih)\*P(Suara Sah Partai | Tidak Terpilih)\*P(Daerah Pemilihan | Tidak Terpilih)\*P(Keanggotaan | Tidak Terpilih)

P(X|Tidak Terpilih) = 0,0500000\*0,3000000\*0,0031048\*0,0021037\*0,0000000\*0,5500000

P(X| Tidak Terpilih) = 0,0000000000000

Tabel 7. Probabilitas Akhir Setiap Kelas.

Table with 3 columns: NAMA, TERPILIH, TIDAK TERPILIH. Rows list names like AGUS EFENDI, H. MUHAMMAD RIZAL, SH, M.Si, etc., with their respective probability values.

f. Probabilitas Akhir

Probabilitas akhir didapat setelah proses perhitungan nilai probabilitas akhir kelas selesai dilakukan. Berikut adalah proses perhitungan probabilitas akhir untuk setiap kelas dari masing-masing alternatif.

Perhitungan probabilitas akhir adalah sebagai berikut:

P(Terpilih|X) = P(Terpilih)\*P(X)

P(Terpilih|X) = 0,459459\*0,0000000027341

P(Terpilih|X) = 0,000000001256225

P(Tidak Terpilih|X) = P(Tidak Terpilih)\*P(X)

P(Tidak Terpilih|X) = 0,540541\*0,0000000000000

P(Tidak Terpilih|X) = 0,000000000000000

Penyelesaian diatas menggunakan satu sampel data uji yaitu alternatif. Probabilitas akhiuntuk masing-masing alternatif, dapat dilihat pada



**Tabel 8.** Probabilitas Akhir.

NAMA	TERPILIH	TIDAK TERPILIH
AGUS EFENDI	0,0000000012562	0,0000000000000
H. MUHAMMAD RIZAL, SH, M.Si	0,0000000088497	0,0000000001512
GOUSTA FERIZA, S.H., M.H	0,0000000014128	0,0000000000000
KOKOK H. DIRGANTORO	0,0000000000000	0,0000000029618
ARI GUNAWAN	0,0000000000000	0,0000000054586
DESY YUSANDI S.E	0,0000000008412	0,0000000000000
AKLINA NOVA RIKA S.IP	0,0000000000000	0,0000000038427
DWI UTAMI, SH	0,0000000000000	0,0000000078964
DR. IR. SYLVIA WISNA DARWIS, SE., M.M.	0,0000000000000	0,0000000084427
ZULHIDJAYANTO HUSIN	0,0000000000000	0,0000000000000
HJ. MARYANA SE	0,0000000000000	0,0000000105859
ASTRID OCTAVIN SARMARIA	0,0000000000000	0,0000000134734
DRS. MUKLIS MM	0,0000000000000	0,0000000000000

Hasil akurasi menggunakan perhitungan manual

CLASS		
PREDICTION	TERPILIH	TIDAK TERPILIH
TERPILIH	3	1
TIDAK TERPILIH	0	9


  

ACCURACY
92,31%

**Gambar 1.** Hasil akurasi perhitungan manual.

### 3.1.3 Pengujian menggunakan Rapidminer

Hasil pengujian Model Algoritma Naive Bayes Classifier menggunakan aRapidminer untuk perbandingan keduanya, dapat dilihat pada gambar berikut:



	true TERPILIH	true TIDAK TERPILIH	class precision
pred TERPILIH	3	1	75.00%
pred TIDAK TERPILIH	0	9	100.00%
class recall	100.00%	90.00%	

**Gambar 2.** Hasil akurasi Rapidminer.

Keterangan :

1. Jumlah prediksi Terpilih dan ternyata benar Terpilih adalah 3 record.
2. Jumlah prediksi Terpilih dan ternyata benar Tidak Terpilih adalah 1 record.



3. Jumlah prediksi Tidak Terpilih dan ternyata benar Tidak Terpilih adalah 9 record.
4. Jumlah prediksi Tidak Terpilih dan ternyata benar Terpilih adalah 0 record.

Nilai Accuracy sebesar 92,31%. Class precision pada prediksi Terpilih memiliki nilai 75%, sedangkan pada prediksi Tidak Terpilih memiliki nilai 100%. Class recall pada true Terpilih memiliki nilai 100%, sedangkan pada true Tidak Terpilih memiliki nilai 90%.

## 4. IMPLEMENTASI

### 4.1 Implementasi Algoritma Naive Bayes dengan Python

#### a. Menyiapkan Data dan Library

Menyiapkan Library yang diperlukan pada Python untuk menunjang pengolahan data.

```
import numpy as np
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score, classification_report
import plotly.figure_factory as ff
import plotly.express as px
import matplotlib.pyplot as plt
```

[3] ✓ 4.7s

**Gambar 3.** Library Python.

Seperti yang dapat dilihat beberapa Library yang digunakan ialah Numpy, Pandas, Matplotlib, Plotly dan juga Sklearn. Library tersebut nantinya yang akan berguna untuk membantu pengolahan data mining algoritma naive bayes.

#### b. Import Data

Jika sudah menyiapkan Library dan menjalankannya, selanjutnya ialah melakukan import data yang ingin diolah ke dalam Python, data tersebut dapat berbentuk file Excel maupun CSV.

```
df = pd.read_csv('data.csv')
```

NO	NO PASI	PRODI	FAKTA	PELTI	SIKAP	LEMB. BELAJAR	JUARA SAH. OLIMPI	JUARA SAH. PINTA	SIKAP. PERSEKUTAN	KEBERSIHAN	REKOR
1	1	1	SAKUN	SAKUN	1	1000	1000	1000	1000	1000	1000
2	2	2	SAKUN	SAKUN	2	1000	1000	1000	1000	1000	1000
3	3	3	SAKUN	SAKUN	3	1000	1000	1000	1000	1000	1000

**Gambar 4.** Proses Import Data.

Script di atas digunakan untuk mengimpor dataset langsung dari file untuk ditampilkan pada Python. Maka tampilan dataframe dapat terlihat seperti tampak pada gambar di atas. Setelah dataset berhasil terimpor, selanjutnya hal yang dapat dilakukan pada dataset tersebut adalah mendapatkan informasi terkait jumlah kolom, tipe data, jumlah seluruh data yang terdapat pada dataset.

### c. Pemilihan Model

Pada penelitian data mining metode yang digunakan adalah metode naive bayes, maka model yang digunakan juga harus sesuai dengan penelitian yang diharapkan. Selanjutnya untuk menggunakan model naive bayes dengan gaussian maka harus dilakukana pemanggilan model dengan cara mengimpornya, untuk melakukannya dapat menggunakan script sebagai berikut:

```
from sklearn.naive_bayes import GaussianNB
✓ 0.0s

model = GaussianNB()
✓ 0.0s
```

**Gambar 5.** Pemilihan Model.

### d. Melakukan pembagian data training dan testing

Setelah variabel data independen dan variabel dependen sudah ditentukan, agar analisis menggunakan klasifikasi Naive Bayes dapat dilakukan, selanjutnya dilakukan pembagian dataset menjadi data training dan juga data testing. Untuk melakukannya dapat menggunakan script, sebagai berikut:

```
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.3, random_state=1)
180 ✓ 0.0s

x_train.shape, x_test.shape
181 ✓ 0.0s
182 ((944, 6), (485, 6))
```

**Gambar 6.** Pembagian Dataset.

Split data dilakukan untuk membagi data menjadi dua bagian, yaitu data training dan data testing di mana pada script di atas proporsi yang digunakan pada penelitian ini membagi dataset menjadi 70% training set dan 30% test set. Yang artinya dari 1349 data, pada training set akan berisi 944 data dan test set berisi 405 data.

### e. Pelatihan Model

Pada bagian ini dilakukan pelatihan model algoritma pada data yang akan diuji, script yang digunakan adalah, sebagai berikut:

```
model = GaussianNB()
model.fit(x_train, y_train)
✓ 0.0s

GaussianNB
GaussianNB()
```

**Gambar 7.** Pelatihan Model.

### f. Classification Report

Berikut adalah hasil implementasi metode algoritma naive bayes untuk memprediksi anggota legislatif terpilih daerah pemilihan provinsi banten, sebagai berikut :

```
print(classification_report(y_test, y_pred))

accuracy = accuracy_score(y_test, y_pred)
print("Akurasi:", accuracy)
```

✓ 0.0s

	precision	recall	f1-score	support
0	0.97	0.98	0.97	379
1	0.65	0.50	0.57	26
accuracy			0.95	405
macro avg	0.81	0.74	0.77	405
weighted avg	0.95	0.95	0.95	405

Akurasi: 0.9506172839506173

**Gambar 8.** Classification Report.

Pengecekan Akurasi pada model naive bayes diperoleh sebesar 95%, yang artinya dari seluruh melakukan prediksi adalah sebesar 95%. Nilai akurasi yang didapat dianggap sudah baik dalam melakukan prediksi dan dapat dijadikan sebagai acuan untuk selanjutnya dapat dikembangkan menjadi sebuah informasi baru untuk para calon anggota legislatif maupun partai pengusung untuk melakukan prediksi kemungkinan terpilihnya calon anggota legislatif. Hal ini karena berdasarkan nilai aktual dan juga prediksi yang diberikan oleh algoritma naive bayes yang terlihat dalam melakukan prediksi tidak banyak melakukan kesalahan sebagaimana tampak pada gambar.

## 5. KESIMPULAN

Berdasarkan hasil analisa dan penelitian yang telah dilakukan dalam memprediksi anggota legislatif terpilih daerah pemilihan provinsi banten tahun 2019, menggunakan data mining dengan algoritma Naive Bayes, maka dapat disimpulkan bahwa Algoritma Naive Bayes terbukti dapat digunakan untuk melakukan prediksi anggota legislatif terpilih daerah pemilihan provinsi banten tahun 2019, sehingga dapat menghasilkan informasi baru. Serta tingkat akurasi yang dihasilkan menggunakan metode algoritma Naive Bayes mencapai 95% yang merupakan nilai yang sudah baik dalam proses memprediksi menggunakan data mining metode algoritma Naive Bayes.

## REFERENCES

- Ibrahim, M. B., Sari, F. P., Kharisma, L. P. I., Kertati, I., Artawan, P., Sudipa, I. G. I., Simanihuruk, P., Rusmayadi, G., Muhammadiyah, M., Nursanty, E., & Lolang, E. (2023). *METODE PENELITIAN BERBAGAI BIDANG KEILMUAN (Panduan & Referensi)*. PT. Sonpedia Publishing Indonesia.
- M.Hum, P. D. M. S. S., S. H. ., MCL MPA ., Dr Imam Kuswahyono, S. H. (2020). *Dinamika Omnibus Law di Era New Normal: Peluang serta Tantangan Bagi Profesi Hukum: PROSIDING SEMINAR NASIONAL*. Media Nusa Creative (MNC Publishing).
- Ph.D, P. D. S., MBA, F. A. A., S. Farm ., Apt ., MSM, & Najib, P. D. M. (2023). *METODE RISET BISNIS: DASAR-DASAR MENDESAIN DAN MELAKUKAN RISET DI KONTEKS BISNIS*. Uwais Inspirasi Indonesia.
- Ramadhan, M. A., & Andarsyah, R. (2022). *KLASIFIKASI TEXT SPAM MENGGUNAKAN METODE SUPPORT VECTOR MACHINE DAN NAÏVE BAYES*. Penerbit Buku Pedia.
- Sanubari, T., Prianto, C., & Riza, N. (2020). *Odol (one desa one product unggulan online) penerapan metode Naive Bayes pada pengembangan aplikasi e-commerce menggunakan Codeigniter*. Kreatif.
- Siregar, F., Kartini, D., Mellaz, A., Wulandari, L., Fernandes, A., Wahyu, Y., Hurriyah, H., Dwiputrianti, S., Sandi, I., Putra, I., Riza, F., Antariksa, N., & Ahmad, S. (2019). *Serial Evaluasi Penyelenggaraan Pemilu Serentak 2019 Perihal Penyelenggaraan Kampanye*.
- Suntoro, J. (2019). *Data Mining: Algoritma dan Implementasi dengan Pemrograman PHP*. Elex Media Komputindo.