

ANALISIS FAKTOR YANG MEMPENGARUHI PEMILIHAN GUBERNUR DAERAH KHUSUS JAKARTA MENGGUNAKAN ALGORITMA NAIVE BAYES DAN REGRESI LOGISTIK

¹⁾Yoga Nanda Khoiril Umat, ²⁾Dhiaz Rusyda Nafsyi, ³⁾Diana Kusumaningsih, ⁴⁾Lukman Hakim

^{1,2,3,4)}Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana

^{1,2,3,4)}Jl. Meruya Selatan No. 1 Kembangan Jakarta Pusat – Jakarta - Indonesia

E-mail : 41521010152@student.mercubuana.ac.id, 41521010163@student.mercubuana.ac.id,
41521010124@student.mercubuana.ac.id, lukman_hakim@mercubuana.ac.id

ABSTRAK

Pemilihan Gubernur dan Wakil Gubernur Daerah Khusus Jakarta (DKJ) tahun 2024 melibatkan masyarakat dalam menentukan pemimpin untuk periode 2024-2029. Penelitian ini bertujuan menganalisis faktor-faktor yang mempengaruhi keputusan pemilih menggunakan algoritma Naive Bayes dan Regresi Logistik. Data survei dikumpulkan melalui kuesioner Google Form dari masyarakat Jakarta yang berusia 17-71 tahun. Proses analisis melibatkan beberapa tahapan, yaitu identifikasi masalah, studi literatur, pengumpulan data, preprocessing, serta pembagian data menjadi data latih dan data uji. Algoritma Naive Bayes digunakan untuk memprediksi parameter klasifikasi berdasarkan pendidikan, popularitas, dan rekam jejak calon, sementara Regresi Logistik memprediksi faktor yang mempengaruhi keputusan pemilih. Naive Bayes menunjukkan akurasi tinggi dengan keunggulan dalam kecepatan dan pengolahan data berukuran besar, sementara Regresi Logistik menunjukkan kekuatan dalam analisis klasifikasi biner dan multinomial. Hasil penelitian menunjukkan bahwa faktor rekam jejak memiliki pengaruh signifikan terhadap keputusan pemilih. Akurasi prediksi Naive Bayes mencapai 85,00% dan Regresi Logistik 80,00%. Hasil analisis juga mengungkapkan bahwa faktor popularitas dan pendidikan calon memiliki pengaruh rasional terhadap keputusan pemilih, meskipun tidak sekuat rekam jejak. Selain itu, penggunaan dua algoritma ini memberikan pandangan yang komprehensif dalam memahami perilaku pemilih di Jakarta. Berdasarkan hasil yang ada, dapat disimpulkan bahwa calon gubernur dan wakil gubernur juga bisa berfokus pada peningkatan rekam jejak dan popularitas mereka untuk meningkatkan peluang terpilih.

Kata Kunci: Pemilihan Gubernur, Daerah Khusus Jakarta, Naive Bayes, Regresi Logistik.

ABSTRACT

The election of the Governor and Deputy Governor of the Special Region of Jakarta (DKJ) in 2024 involves the community in determining leaders for the 2024-2029 period. This research analyzes the factors influencing voter decisions using the Naive Bayes algorithm and Logistic Regression. Survey data was collected via a Google Form questionnaire from Jakarta residents aged 17-71. The analysis process involves several stages: problem identification, literature study, data collection, preprocessing, and dividing the data into training and test data. The Naive Bayes algorithm is used to predict classification parameters based on a candidate's education, popularity, and track record, while Logistic Regression predicts factors that influence voter decisions. Naive Bayes shows high accuracy with advantages in speed and processing large data, while Logistic Regression shows strength in binary and multinomial classification analysis. The research results show that the track record factor significantly influences voter decisions. Naive Bayes prediction accuracy reached 85.00% and Logistic Regression 80.00%. The analysis results also reveal that the candidate's popularity and education factors rationally influence voter decisions, although not as strong as the track record. In addition, using these two algorithms provides a comprehensive understanding of voter behavior in Jakarta. Based on these results, governor and deputy governor candidates should also focus on improving their track record and popularity to increase their chances of being elected.

Keyword: Governor Election, Special Region of Jakarta, Naive Bayes, Logistic Regression.

PENDAHULUAN

Di tahun 2024 ini, Daerah Khusus Jakarta (DKJ) akan menyelenggarakan pesta demokrasi yang sering dikenal sebagai

Pemilihan Kepala Daerah (Pilkada). Masyarakat Jakarta akan memilih calon Gubernur dan Wakil Gubernur untuk memimpin Daerah Khusus Jakarta (DKJ) selama lima tahun kedepan yaitu untuk periode

2024 – 2029. Pilkada serentak akan dilaksanakan pada tanggal 27 November, melibatkan DKI Jakarta beserta 38 provinsi lainnya dan lebih dari 514 kabupaten/kota. Berdasarkan laporan dari Ketua KPU DKI Jakarta, Wahyu Dinata, yang disampaikan pada acara sosialisasi di Hotel Borobudur, Jakarta Pusat, pada tanggal 2 April 2024[1].

Masyarakat seringkali mengalami kebingungan dalam menentukan pilihan berdasarkan faktor apa saja yang seharusnya diutamakan dalam memilih untuk calon Gubernur dan Wakil Gubernur. Sehingga, dalam menghadapi Pemilihan Kepala Daerah (Pilkada) 2024 di

Daerah Jakarta, ada beberapa faktor yang harus diperhatikan oleh Masyarakat dalam membuat keputusan.

Menurut penelitian yang dilakukan oleh Arif Senja Fitriani dalam studinya yang berjudul "Penerapan Data Mining Menggunakan Metode Klasifikasi Naïve Bayes untuk Memprediksi Partisipasi Pemilihan Gubernur", disimpulkan bahwa prediksi partisipasi pemilu berdasarkan dataset yang terdiri dari 300 data yang dibagi menjadi dua bagian yaitu 56% (195 data) sebagai data Training dan 35% (105 data) sebagai data Testing. Hasil prediksi berdasarkan atribut kehadiran menunjukkan bahwa dari 105 data Testing, tingkat akurasi prediksi mencapai 97%, sedangkan tingkat kesalahan prediksi adalah 3%. [2].

Penelitian Mhd. Gading Sadewo et al. membuktikan bahwa Algoritma Naïve Bayes dapat memprediksi kepuasan nasabah Bank BTN KC Pematangsiantar dengan tepat. Algoritma ini menghasilkan tingkat akurasi klasifikasi yang tinggi, mencapai 88%, menunjukkan hasil yang sangat baik. Penelitian ini menggunakan data dari 25 nasabah Bank BTN KC Pematangsiantar, yang dikategorikan menjadi dua kelas: Puas dan Tidak Puas. Perhitungan dengan Algoritma Naïve Bayes

menunjukkan bahwa 20 nasabah masuk kategori Puas dan 5 nasabah masuk kategori Tidak Puas. Pengujian data menggunakan RapidMiner 5.3 dengan algoritma Naïve Bayes menghasilkan akurasi klasifikasi dua kelas sebesar 88%. [3].

Penelitian lain yang dilakukan oleh Bister Purba dan Rian Syahputra dalam jurnal berjudul "Implementasi Metode Naive Bayes Classifier pada Evaluasi Kepuasan Mahasiswa terhadap Pembelajaran Daring" menunjukkan hasil yang menarik. Dari pengujian yang dilakukan terhadap data training yang terdiri dari 97 responden dan data testing sebanyak 12 responden, diperoleh hasil prediksi sebagai berikut yaitu 8 responden diprediksi puas dan ternyata benar-benar puas, 0 responden diprediksi tidak puas tetapi sebenarnya puas, 2 responden diprediksi puas tetapi sebenarnya tidak puas, dan 2 responden diprediksi tidak puas dan ternyata benar-benar tidak puas. Tingkat akurasi keseluruhan mencapai 83,33%. Pada prediksi tidak puas, nilai precision adalah 50%, sementara pada prediksi puas nilai precision mencapai 100%. Selain itu, nilai recall untuk kelas benar-benar tidak puas adalah 50%, sedangkan nilai recall untuk kelas benar-benar puas mencapai 100%. [4].

Penelitian Hendriyana, Ichwanul Muslim Karo Karo, dan Sri Dewi untuk memprediksi potensi donor darah menemukan bahwa ketiga algoritma klasifikasi yang diuji, yaitu Naïve Bayes, Regresi Logistik, dan SVM, menunjukkan performa yang cukup baik. Perbedaan karakteristik masing-masing algoritma menghasilkan variasi performa. Hasil pengujian menunjukkan bahwa Naïve Bayes memiliki akurasi tertinggi (89.90%), diikuti oleh SVM (94.79%) dan Regresi Logistik (82.59%). Performa ini dibandingkan dengan penelitian lain untuk menunjukkan keunggulan Naïve Bayes dalam konteks penelitian ini. [5]

Naive Bayes memiliki keunggulan dalam

mengolah data untuk memprediksi parameter yang diperlukan dalam klasifikasi[6]. Selain itu, algoritma ini juga dikenal dengan kecepatan dan tingkat akurasi yang tinggi saat diterapkan pada basis data berukuran besar. Analisis faktor yang mempengaruhi pemilihan calon Gubernur dan Wakil Gubernur Daerah Khusus Jakarta Periode 2024 - 2029 ini bertujuan untuk mengevaluasi sejauh mana pengaruh berbagai pertanyaan yang diajukan terhadap keputusan masyarakat dalam membuat keputusan.

Sedangkan Algoritma Regresi Logistik memiliki peluang untuk memprediksi faktor mana saja yang dapat mempengaruhi keputusan pemilih dalam memilih Calon Gubernur dan Wakil Gubernur Daerah Khusus Jakarta Periode 2024 - 2029. Algoritma ini sangat sesuai untuk digunakan karena mampu menangani variabel independen yang bersifat biner maupun kontinu, serta dapat memberikan gambaran yang jelas mengenai hubungan antara variabel-variabel tersebut. Urgensi penelitian ini terletak pada pentingnya memahami faktor-faktor yang mempengaruhi pilihan pemilih Calon Gubernur dan Wakil Gubernur Daerah Khusus Jakarta Periode 2024 - 2029 agar strategi kampanye dan kebijakan publik dapat disusun secara lebih efektif dan sesuai dengan kebutuhan masyarakat. Dengan menggunakan Algoritma Regresi Logistik, diharapkan penelitian ini dapat memberikan wawasan mendalam mengenai pengaruh berbagai variabel terhadap keputusan pemilih, sehingga mampu meningkatkan kualitas demokrasi dan partisipasi politik di Jakarta.

METODE

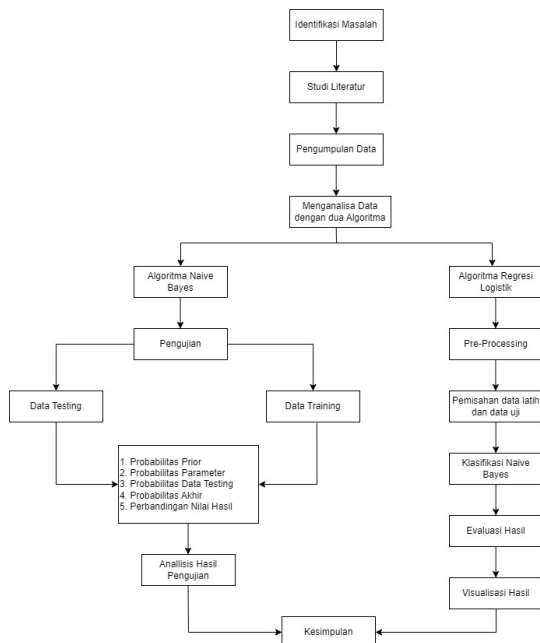
Data survei yang digunakan dalam analisis ini mencakup pertanyaan terkait pendidikan sebelumnya dari calon Gubernur dan Wakil Gubernur, rekam jejak serta popularitasnya. Langkah-langkah yang dilakukan untuk

menganalisis Keputusan pemilih adalah :

- 1) Identifikasi Masalah : Langkah awal dalam penelitian yang bertujuan untuk menentukan isu atau fenomena yang ingin dipelajari atau dipecahkan. Pada tahap ini, dirumuskan tujuan penelitian dan mengidentifikasi pertanyaan atau masalah spesifik yang perlu dijawab.
- 2) Studi Literatur : Langkah kedua adalah proses meninjau Penelitian sebelumnya yang relevan dengan topik yang diambil, serta mengidentifikasi masalah yang sudah ada. Melalui tahapan ini, dapat ditemukan solusi dari permasalahan tersebut.
- 3) Pengumpulan Data : Data yang sudah ada diperoleh dari Formulir Google Form yang berisi pertanyaan sesuai dengan topik yang diambil. Kuisisioner tersebut diisi oleh responden dan Google Form ini memungkinkan untuk menjangkau Masyarakat Jakarta yang berusia 17 – 71 tahun.
- 4) Menganalisis Data dengan dua Algoritma : Data yang sudah terkumpul selanjutnya akan dianalisis dengan algoritma Naive Bayes dan Algoritma Regresi Logistik.
- 5) Pre-Processing : Data yang terdapat dalam kuisisioner google form masih terdapat informasi mengenai Timestamp dan Nomor Telepon. Sehingga perlu dilakukan pembersihan data pada data yang tidak diperlukan.
- 6) Pemisahan Data Latih dan Data Uji: Sebelum memulai proses pelatihan model pembelajaran mesin, dataset akan dibagi menjadi dua bagian, yaitu data latih dan data uji. Data uji digunakan untuk mengevaluasi kinerja algoritma klasifikasi. Dalam tahap analisis ini, 80% dari dataset akan digunakan sebagai data latih, sementara

20% sisanya akan digunakan sebagai data uji.

Gambar 1 berikut menunjukkan tahapan penelitian yang dilakukan:



Gambar 1. Rangkaian Tahapan Penelitian

Regresi Logistik

Regresi logistik adalah salah satu algoritma pembelajaran mesin yang handal digunakan untuk klasifikasi data dengan target bertipe kategori[7]. Regresi Logistik sangat bermanfaat untuk masalah klasifikasi biner. Namun, Regresi Logistik Multinomial dapat digunakan untuk klasifikasi yang mencakup lebih dari dua kategori. Ini memungkinkan kita untuk memprediksi kategori yang bersifat nominal atau ordinal yang terdiri dari lebih dari dua kelas. Kelemahan dari Regresi logistik adalah rentan terhadap underfitting dataset yang kelasnya tidak seimbang, sehingga menghasilkan akurasi yang rendah[8].

Analisis regresi logistik multinomial adalah metode statistik untuk menganalisis hubungan antara variabel dependen dan variabel independen dimana variabel dependen adalah kategoris dan ada dua atau lebih kategori[9].

Bentuk umum fungsi logit dengan variabel respon mempunyai tiga kategori sebagai berikut[10]:

$$g_j(x) = \beta_{j0} + \beta_{j1}x_1 + \beta_{j2}x_2 + \dots + \beta_{jp}x_p$$

Suatu variabel respon dengan tiga kategori akan membentuk dua persamaan logit, di mana masing-masing persamaan ini membentuk regresi logistik multinomial yang membandingkan suatu kelompok kategori terhadap pembandingan[10].

Algoritma Naïve Bayes

Naïve Bayes adalah proses klasifikasi yang sederhana untuk menghitung peluang dengan menggunakan rumus bayes yang digabungkan dengan nilai data yang ada. Dalam penggabungan atau klasifikasi ini, tujuannya adalah untuk memprediksi dan memperkirakan data label dari suatu data uji berdasarkan atribut dan ciri-khasnya sendiri. Seperti contoh, berdasarkan pada data mengenai usia para responden, kegemaran dalam memilih makanan, dan gaya hidup yang berbeda-beda[11]. Naive Bayes adalah statistik klasifikasi yang dapat digunakan untuk memperkirakan kemungkinan setiap kelas tertentu akan memiliki anggota[12].

Naïve Bayes membuat dan menilai model dengan sangat cepat dan skala secara linier dalam jumlah prediksi dan baris[13]. Metode ini perlu menghitung probabilitas setiap nilai atribut untuk membangun model Naive Bayes. Metode ini merupakan metode klasifikasi yang sederhana dan setiap atributnya memiliki sifat yang saling bebas sehingga setiap atributnya dimungkinkan untuk memiliki kontribusi terhadap keputusan akhir [14].

Kelebihan metode ini ialah[15] :

1. Membenahi kuantitatif serta diskrit
2. Tangguh untuk titik kebisingan terisolasi, seperti yang dirata-rata saat mengevaluasi

probabilitas bersyarat data.

3. Hanya dengan jumlah kecil data training untuk memperkirakan parameter (rata – rata serta variansi yang berasal dari variabel) untuk klasifikasi.
4. Skala nilai yang hilang disebabkan oleh pengabaian kejadian selama estimasi kemungkinan peluang.
5. Cepat serta menghemat waktu
6. Kuat terhadap karakteristik yang tidak relevan.

Kekurangan Metode Naïve Bayes[15]:

1. Memperkirakan bahwa karakteristiknya bebas.
2. Ini tidak berlaku, Jika probabilitas bersyarat adalah 0, probabilitas prediksi juga akan menjadi 0.

Untuk menyelesaikan metode Naïve Bayes, dapat dilakukan dengan menggunakan persamaan-persamaan berikut[2]:

- 1) Membaca data training
- 2) Menghitung jumlah dan Probabilitas Prior

$$P(C_i) = \frac{S_i}{s} \quad (1)$$

- 3) Menghitung Probabilitas Training

$$P(K|D) = \frac{P(D|K) \cdot P(K)}{P(D)} \quad (2)$$

Dimana:

D : Data kelas yang belum diketahui.

K : Hipotesis suatu class yang spesifik

$P(K|D)$: Probabilitas hipotesis K berdasarkan kondisi D

$P(K)$: Probabilitas sesuai dengan hipotesis D.

$P(D|K)$: Probabilitas hipotesis D berdasarkan kondisi K.

$P(D)$: Probabilitas sesuai dengan

hipotesis K

$$4) \text{ Hitung Nilai Probabilitas Data Testing} \\ P(K|a_1, \dots, a_n) = P(D) * P(a_1, \dots, a_n|D) \quad (3)$$

$$5) \text{ Menghitung Probabilitas Akhir} \\ P(K|D) = P(D|K) * P(K|D) \quad (4)$$

$$6) \text{ Menghitung Nilai Akhir} \\ C = \arg \max P(D|K) \quad (5)$$

HASIL DAN PEMBAHASAN

Analisa dengan naïve bayes

Dalam analisis data, proses yang dilakukan melibatkan pengkategorian setiap pertanyaan dalam kuesioner dengan label A1 hingga A16. Ada empat parameter yang digunakan, dan masing-masing parameter memiliki 16 kategori, seperti yang ditunjukkan dalam tabel 1 dibawah ini:

Tabel 1. Parameter

Parameter I : Pendidikan

Kategori | Pertanyaan

- A1 Menurut Anda Apakah penting pendidikan formal yang dimiliki oleh calon gubernur yang anda pilih dalam memimpin Jakarta?
- A2 Menurut Anda Apakah calon gubernur Jakarta yang anda pilih harus memiliki pengalaman pendidikan yang terkait dengan masalah-masalah sosial dan ekonomi yang dihadapi oleh masyarakat Jakarta?
- A3 Menurut Anda Apakah calon gubernur yang anda pilih memiliki pendidikan lanjutan seperti gelar master atau doktor lebih mampu menghadapi tantangan kompleks dalam mengelola Jakarta?
- A4 Menurut Anda Apakah calon gubernur Jakarta yang anda pilih harus memiliki latar belakang

pendidikan yang relevan dengan bidang pemerintahan atau manajemen publik?

- A5 Menurut Anda Apakah pendidikan formal calon gubernur yang anda pilih dapat mempengaruhi keputusan Anda untuk memberikan suara pada pemilihan tersebut?

Parameter II : Popularitas

- A6 Menurut Anda Apakah penting popularitas calon gubernur yang anda pilih dalam menentukan kemampuannya untuk memimpin Jakarta?
- A7 Menurut Anda Apakah efektif popularitas calon gubernur yang anda pilih dalam memengaruhi keputusan masyarakat Jakarta?
- A8 Menurut Anda apakah popularitas calon gubernur yang anda pilih cenderung mencerminkan kualitas yang lebih baik dalam menjalankan tugas-tugasnya?
- A9 Menurut Anda apakah popularitas calon gubernur yang anda pilih dapat menjamin keberhasilannya dalam memecahkan masalah-masalah kritis di Jakarta?
- A10 Menurut Anda Apakah berpengaruh popularitas calon gubernur yang anda pilih terhadap tingkat partisipasi pemilih dalam pemilihan gubernur?

Parameter III: Rekam Jejak

- A11 Menurut Anda apakah pengalaman dan prestasi sebelumnya membuat calon gubernur Jakarta 2024 yang anda pilih menjadi lebih baik?

- A12 Menurut Anda apakah rekam jejak calon gubernur Jakarta yang anda pilih akan mempengaruhi keputusan Anda dalam memilih?

- A13 Menurut Anda Apakah rekam jejak calon gubernur Jakarta 2024 yang anda pilih dalam bidang politik atau pemerintahan dapat menjadi indikator yang baik untuk kinerja mereka di masa depan?

- A14 Apakah Anda merasa bahwa rekam jejak calon gubernur Jakarta 2024 yang anda pilih mampu menghadapi tantangan dan krisis selama masa jabatan sebelumnya dapat menjadi indikator yang baik untuk kemampuan mereka dalam menangani situasi sulit di masa depan?

- A15 Menurut Anda apakah Anda yakin calon gubernur yang anda pilih memiliki kemampuan memimpin Jakarta berdasarkan rekam jejak kinerjanya?

Setiap parameter mencakup beberapa pertanyaan yang diajukan kepada masyarakat dengan menggunakan skala penilaian empat tingkat, yaitu, Sangat Setuju (SS), Setuju (S), Tidak Setuju (TS), dan Sangat Tidak Setuju (STS).

Jumlah data yang digunakan dalam penelitian ini adalah 80 data, dengan 69 data termasuk dalam kategori 'Pengaruh' dan 11 data termasuk dalam kategori 'Tidak Pengaruh'. Untuk pengujian, terdapat 20 data uji yang akan diprediksi, seperti yang ditunjukkan dalam gambar 2 berikut ini:

Nama Anda	Usia	Domisili KTP	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	A15	Apakah pertanyaan kuesioner mempengaruhi pilihan anda?	Nama Bakal Calon Gubernur Yang Dipilih
Budi Santoso	22 - 26 Tahun	Jakarta Barat	3	3	3	3	3	3	3	3	3	3	3	4	4	4	4	Pengaruh	Erwin Aksa
Aditya Pratama	27 - 31 Tahun	Jakarta Barat	2	4	1	3	4	2	4	3	2	3	1	1	2	4	2	Pengaruh	Ridwan Kamil
Rizki Cahyono	32 - 36 Tahun	Jakarta Selatan	2	2	3	4	1	2	3	2	4	3	3	3	2	4	2	Pengaruh	Ahmed Zaki Iskandar
Siti Aisyah	17 - 21 Tahun	Jakarta Pusat	4	1	2	3	1	3	2	3	4	3	4	3	2	3	4	Pengaruh	Ridwan Kamil
Dewi Lestari	37 - 41 Tahun	Jakarta Timur	4	1	1	3	2	1	1	3	1	3	3	3	4	2	3	Pengaruh	Anies Baswedan
Nia Puspita	42 - 46 Tahun	Jakarta Barat	3	2	1	3	4	2	2	3	4	2	3	1	2	3	4	Tidak Pengaruh	Anies Baswedan
Wahyu Nugroho	27 - 31 Tahun	Jakarta Selatan	3	2	3	3	1	1	3	4	2	4	3	1	3	4	2	Pengaruh	Ridwan Kamil
Andika Wijaya	52 - 56 Tahun	Jakarta Selatan	1	3	2	4	2	3	3	2	1	4	2	3	2	4	1	Pengaruh	Ahmad Saroni
Satria Putra	27 - 31 Tahun	Jakarta Seribu	4	2	3	1	3	4	4	4	3	3	4	2	1	3	2	Pengaruh	Mardani Ali Sera
Dwi Hartono	32 - 36 Tahun	Jakarta Barat	3	1	2	3	2	2	2	2	1	1	3	3	1	2	2	Pengaruh	Anies Baswedan
Rina Fitriani	42 - 46 Tahun	Jakarta Pusat	3	2	3	4	3	4	4	4	4	3	3	2	2	4	2	Pengaruh	Erwin Aksa
Maya Sari	22 - 26 Tahun	Jakarta Utara	4	3	3	2	2	3	4	1	3	1	3	2	4	3	2	Pengaruh	Ahmad Saroni
Putri Amelia	22 - 26 Tahun	Jakarta Selatan	4	2	2	3	4	3	3	2	2	3	4	2	1	3	2	Pengaruh	Anies Baswedan
Rizka Dewi	27 - 31 Tahun	Jakarta Barat	4	1	3	2	4	1	2	2	3	1	4	2	3	3	1	Pengaruh	Ridwan Kamil
Fajar Rahman	52 - 56 Tahun	Jakarta Selatan	4	2	3	3	2	2	4	3	4	2	4	2	4	1	3	Tidak Pengaruh	Erwin Aksa
Bayu Seliawan	22 - 26 Tahun	Jakarta Timur	3	2	3	1	2	2	1	3	4	2	3	1	1	2	4	Pengaruh	Mardani Ali Sera
Kurniawan	17 - 21 Tahun	Jakarta Barat	4	1	2	2	2	1	3	1	2	4	1	2	3	2	3	Pengaruh	Ridwan Kamil
Rendra Wibowo	22 - 26 Tahun	Jakarta Barat	4	2	3	2	1	3	2	3	1	4	3	3	2	2	3	Pengaruh	Anies Baswedan
Denny Saputra	32 - 36 Tahun	Jakarta Utara	3	2	1	3	4	2	2	1	3	2	2	3	4	4	3	Pengaruh	Ahmad Saroni
Dina Nurul	32 - 36 Tahun	Jakarta Pusat	4	1	1	1	3	2	2	2	3	1	3	4	3	4	4	Pengaruh	Ridwan Kamil

Gambar 2. Data Testing

- Menghitung jumlah Probabilitas Prior
Jumlah data yang digunakan dalam penelitian ini adalah sebanyak 80 data, dengan 69 data termasuk dalam kategori 'Pengaruh' dan 11 data dalam kategori 'Tidak Pengaruh'. Langkah-langkah menghitung nilai probabilitas responden adalah sebagai berikut.:

$$P(R|P) = \frac{69}{80} = 0,8625$$

$$P(R|TP) = \frac{11}{80} = 0,1375$$

- Menghitung Probabilitas pada masing-masing parameter.

Untuk parameter I hingga parameter III, dan untuk kategori A1 hingga A15, perhitungan probabilitas parameter dilakukan.

- Parameter 1 : Pendidikan (Kategori A1)

$$P(SS|P) = \frac{19}{69} = 0,275$$

$$P(SS|TP) = \frac{1}{11} = 0,090$$

$$P(S|P) = \frac{32}{69} = 0,463$$

$$P(S|TP) = \frac{8}{11} = 0,727$$

$$P(TS|P) = \frac{15}{69} = 0,217$$

$$P(TS|TP) = \frac{1}{11} = 0,090$$

$$P(STS|P) = \frac{3}{69} = 0,043$$

$$P(STS|TP) = \frac{1}{11} = 0,090$$

Tabel 2: Probabilitas Parameter I Kategori A1

Prob. 1	Pengaruh	Tidak Pengaruh	Prob. P	Prob. TP
SS	19	1	0,275	0,090
S	32	8	0,463	0,727
TS	15	1	0,217	0,090
STS	3	1	0,043	0,090
Jumlah	69	11	1	1

Oleh karena itu, dilakukan proses penghitungan parameter untuk setiap parameter setiap kategori hingga maksimal parameter III Kategori A15 seperti terlihat pada tabel 3 di bawah ini :

Tabel 3: Probabilitas Parameter III Kategori A15

Probabilitas 15	Pengaruh	Tidak Pengaruh	Prob. P	Prob. TP
SS	23	2	0,333	0,181
S	30	8	0,434	0,727
TS	14	0	0,202	0
STS	2	1	0,028	0,090
Jumlah	69	11	1	1

3. Perhitungan Probabilitas untuk Data Uji Data uji diperoleh sebanyak 20 responden, 80 hingga 100 pada Gambar 2, dan data uji probabilitas untuk responden Pengaruh (P) dan Tidak Pengaruh (TP).

$$P(R1|TP) = 0,727 * 0,818 * 0,090 * 0 * 0,636 * 0,545 * 0,272 * 0,818 * 0,636 * 0,636 * 0,272 * 0,727 * 0,636 * 0,181 * 0,181$$

$$P(R1|TP) = 0,0000000000$$

a) Probabilitas responden terhadap data uji "P" Berdasarkan persamaan 3, perhitungan responden terhadap data uji "P" dapat dilakukan sebagai berikut:

Oleh karena itu, dilakukan perhitungan serupa terhadap data uji "TP" dari responden 2 hingga responden 40.

$$P(R1|P) = P(A1 = S|P) * P(A2 = S|P) * P(A3 = TS|P) * P(A4 = S|P) * P(A5 = S|P) * P(A6 = S|P) * P(A7 = S|P) * P(A8 = S|P) * P(A9 = S|P) * P(A10 = S|P) * P(A11 = S|P) * P(A12 = S|P) * P(A13 = S|P) * P(A14 = S|P) * P(A15 = S|P)$$

c) Menghitung probabilitas akhir responden.

a) Probabilitas "Pengaruh (P)" :

$$P(P|R1) = P(R1|P) * P(K|P) = 0,0000005237 * 0,8625 = 0,0000004517$$

$$P(R1|P) = 0,4580 * 0,375 * 0,2083 * 0,4583 * 0,5000 * 0,2917 * 0,5417 * 0,3750 * 0,5000 * 0,5417 * 0,4167 * 0,2500 * 0,5000 * 0,3750 * 0,3750$$

Selanjutnya, probabilitas "P" dihitung hingga jumlah responden 20 dengan $P(P | R20)$. Hasilnya ditampilkan dalam tabel 4 berikut:

$$P(R1|P) = 0,00000096373$$

Oleh karena itu, dilakukan perhitungan serupa terhadap data uji "P" dari responden 2 hingga responden 40.

b) Probabilitas responden data testing "TP"

$$P(R1|TP) = P(A1 = S|TP) * P(A2 = S|TP) * P(A3 = SS|TP) * P(A4 = SS|TP) * P(A5 = S|TP) * P(A6 = S|TP) * P(A7 = SS|TP) * P(A8 = S|TP) * P(A9 = S|TP) * P(A10 = S|TP) * P(A11 = SS|TP) * P(A12 = S|TP) * P(A13 = S|TP) * P(A14 = SS|TP) * P(A15 = SS|TP)$$

Tabel 4. Nilai Probabilitas Pengaruh

R	P(R P)	P(P R)
1	0,0000005237	0,0000004517
2	0,0000006911	0,0000005961
3	0,0000007925	0,0000006836
4	0,0000004854	0,0000004186
5	0,0000000053	0,0000000046
6	0,0000001771	0,0000001527
7	0,0000000128	0,0000000110
8	0,0000000031	0,0000000027
9	0,0000000000	0,0000000000
10	0,0000000015	0,0000000013
11	0,0000000022	0,0000000019
12	0,0000000021	0,0000000018
13	0,0000000017	0,0000000015
14	0,0000000020	0,0000000017
15	0,0000000000	0,0000000000

16	0,0000000069	0,0000000059
17	0,0000000212	0,0000000183
18	0,0000000004	0,0000000003
19	0,0000000001	0,0000000001
20	0,0000000001	0,0000000001

b) Probabilitas “Tidak Pengaruh (TP)”

$$\begin{aligned}
 &: \\
 P(TP|R1) &= P(R1|TP) * P(K|TP) \\
 &= 0,0000000000 * \\
 &0,1375 \\
 &= 0,0000000000
 \end{aligned}$$

Selanjutnya, probabilitas "TP" dihitung hingga jumlah responden 20 dengan P(TP|R20). Hasilnya ditampilkan dalam tabel 5 berikut:

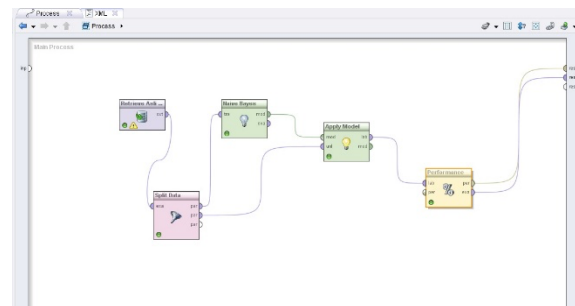
Tabel 5. Nilai Probabilitas Tidak Pengaruh

R	P(R TP)	P(TP R)
1	0,0000000000	0,0000000000
2	0,0000000000	0,0000000000
3	0,0000011804	0,0000001623
4	0,0000000000	0,0000000000
5	0,0000000000	0,0000000000
6	0,0000000000	0,0000000000
7	0,0000000000	0,0000000000
8	0,0000000000	0,0000000000
9	0,0000000000	0,0000000000
10	0,0000000000	0,0000000000
11	0,0000000000	0,0000000000
12	0,0000000000	0,0000000000
13	0,0000000000	0,0000000000
14	0,0000000000	0,0000000000
15	0,0000000000	0,0000000000
16	0,0000000000	0,0000000000
17	0,0000000000	0,0000000000
18	0,0000000000	0,0000000000
19	0,0000000000	0,0000000000
20	0,0000000000	0,0000000000

Analisa Naïve Bayes Pada RapidMiner

Setelah analisis manual, kami melakukan analisis manual menggunakan metode Naïve Bayes. Langkah selanjutnya adalah melakukan pengujian pada menggunakan tool Rapidminer sebagai berikut:

- 1) Gambar 3 berikut ini menunjukkan pemodelan Algoritma Naïve Bayes menggunakan RapidMiner:



Gambar 3. Pemodelan Algoritma Naïve Bayes dengan RapidMiner.

- 2) Accuracy
Gambar 4 berikut ini menunjukkan tingkat akurasi:

Multiclass Classification Performance			
accuracy: 85.00%			
	True Pengaruh	True Total Pengaruh	class precision
pred: Pengaruh	15	1	93.75%
pred: Tidak Pengaruh	2	2	50.00%
class recall	86.24%	56.07%	

Gambar 4. Nilai Accuracy Algoritma Naïve Bayes dengan RapidMiner.

Gambar tersebut menampilkan performa sebuah model klasifikasi dengan tingkat akurasi 85.00%. Dalam hal ini, model berhasil mengklasifikasikan 85% data uji dengan benar, baik untuk kategori "Pengaruh" maupun "Tidak Pengaruh".

Analisa Regresi Logistik

Setelah melakukan analisis dengan Algoritma Naive Bayes, kami melakukan analisis menggunakan Algoritma Regresi Logistik dengan bahasa pemrograman Python di Google Collab dan menggunakan Library

seperti Gambar 5 dibawah ini :

```
[ ] import pandas as pd
from sklearn.preprocessing import LabelEncoder, StandardScaler
from sklearn.linear_model import LogisticRegression
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score, classification_report
```

Gambar 5. Library python algoritma regresi logistik

Arti dari library diatas adalah :

1. 'Pandas' : Library utama untuk menganalisis suatu data di Python.
2. 'LabelEncoder' : Mengubah label ataupun atribut variabel bersifat kategorikal menjadi nilai numerik.
3. 'StandardScaler' : Menstandarkan atribut dengan menghilangkan rata-rata ke skala varian satuan.
4. 'LogisticRegression' : Kelas yang mengimplementasikan regresi logistik dan digunakan untuk klasifikasi data biner (0 dan 1) atau multinomial serta digunakan untuk memprediksi probabilitas.
5. 'train_test_split' : Untuk membagi data menjadi dua subset. Satu untuk pelatihan model dan pengujian model.
6. 'Accuracy_score' : Untuk menghitung akurasi dari model klasifikasi.
7. 'Classification_report' : Untuk menghasilkan laporan yang menunjukkan metrik utama dalam evaluasi kinerja klasifikasi.

```
[ ] # Definisikan faktor-faktor
factors = {
    'Faktor Pendidikan': ['A1', 'A2', 'A3', 'A4', 'A5'],
    'Faktor Popularitas': ['A6', 'A7', 'A8', 'A9', 'A10'],
    'Faktor Rekam Jejak': ['A11', 'A12', 'A13', 'A14', 'A15']
}
```

Gambar 6. Pengelompokan Pertanyaan berdasarkan Parameter.

Dalam pengelompokan faktor-faktor yang mempengaruhi berbagai aspek, kita dapat

membaginya menjadi beberapa parameter utama. Parameter pertama adalah Faktor Pendidikan, yang mencakup lima pertanyaan yang diberi kode unik dari A1 hingga A5. Pertanyaan-pertanyaan ini berhubungan dengan aspek-aspek yang berkaitan dengan pendidikan calon gubernur dan wakil gubernur, seperti tingkat pendidikan yang telah dicapai, kualitas institusi pendidikan, prestasi akademik, keahlian khusus yang diperoleh, dan pengalaman belajar.

Selanjutnya, ada Faktor Popularitas, yang diwakili oleh kode A6 hingga A10. Pertanyaan-pertanyaan dalam faktor ini berkaitan dengan popularitas atau pengaruh seseorang di masyarakat. Hal ini bisa mencakup jumlah pengikut di media sosial, citra publik, frekuensi muncul di media, keterlibatan dalam kegiatan sosial atau komunitas, serta pengakuan atau penghargaan yang diterima.

Parameter ketiga adalah Faktor Rekam Jejak, yang terdiri dari pertanyaan-pertanyaan dengan kode A11 hingga A15. Pertanyaan-pertanyaan ini mencakup aspek-aspek yang berkaitan dengan rekam jejak atau sejarah seseorang dalam berbagai bidang. Ini bisa mencakup pengalaman kerja, proyek atau inisiatif yang telah dijalankan, kontribusi di bidang tertentu, reputasi profesional, serta catatan pencapaian atau kegagalan di masa lalu.

```
[ ] # Split data menjadi training dan testing set dengan rasio 80/20
train_df, test_df = train_test_split(df, test_size=0.2, random_state=42)
```

Gambar 7. Pembagian data dengan ratio 80 / 20.

Berdasarkan gambar 7 diatas, artinya adalah kode tersebut menghitung dan menampilkan akurasi keseluruhan model regresi logistik pada data testing. Pertama, fungsi *accuracy_score* dari library *sklearn.metrics* digunakan untuk menghitung akurasi dengan membandingkan

prediksi model (`model.predict(X_test)`) dengan label sebenarnya (`y_test`). Nilai akurasi yang diperoleh akan dikalikan dengan 100 untuk mengubahnya menjadi persentase, dan hasilnya kemudian dicetak dengan dua angka desimal. Pernyataan print ini memberikan informasi tentang seberapa baik model secara keseluruhan dalam memprediksi label data uji, yang merupakan indikator penting untuk mengevaluasi kinerja model regresi logistik.

```
[ ] # Fit model regresi logistik untuk setiap faktor dan dapatkan probabilitas prediksi
for factor, columns in factors.items():
    X_train = train_df[columns]
    X_test = test_df[columns]
    model = LogisticRegression(max_iter=1000)
    model.fit(X_train, y_train)
    train_score = model.score(X_train, y_train)
    results[factor] = train_score
    predicted_probabilities[factor] = model.predict_proba(X_test[:, 1])
    test_accuracy = accuracy_score(y_test, model.predict(X_test))
    print(f"{factor} - Akurasi data test: {test_accuracy * 100:.2f}%")
```

Gambar 8. Penerapan Bahasa pemrograman Python dengan Algoritma Regresi Logistik.

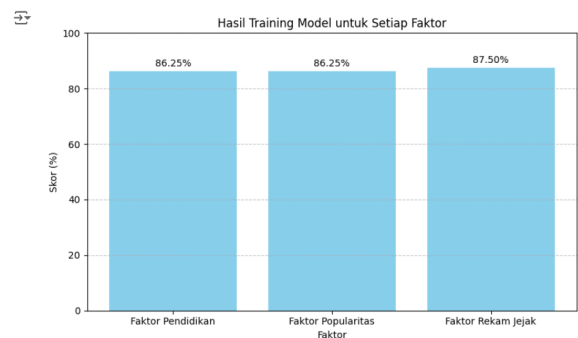
Kode yang dibuat seperti pada gambar 8 diatas untuk melatih dan mengevaluasi model regresi logistik untuk setiap faktor yang ada dalam dataset. Untuk setiap faktor, data pelatihan (`train_df`) dan data pengujian (`test_df`) dipisahkan berdasarkan kolom-kolom yang sesuai. Kemudian, model regresi logistik dengan batas maksimal iterasi sebanyak 1000 dilatih menggunakan data pelatihan (`X_train`) dan label pelatihan (`y_train`). Setelah model dilatih, persentase akurasi pada data pelatihan dihitung dan disimpan dalam variabel `results` dengan nama faktor sebagai kuncinya. Model yang telah dilatih ini kemudian digunakan untuk memprediksi probabilitas kelas Pilihan pada data pengujian (`X_test`), dan probabilitas tersebut disimpan dalam variabel `predicted_probabilities`. Akhirnya, akurasi model pada data pengujian dihitung dan ditampilkan, dengan akurasi yang diperoleh dari perbandingan antara prediksi model dan label pengujian (`y_test`).

```
[ ] # Hitung akurasi keseluruhan pada data test
overall_accuracy = accuracy_score(y_test, model.predict(X_test))
print(f"Akurasi keseluruhan pada data test: {overall_accuracy * 100:.2f}%")
```

Akurasi keseluruhan pada data test: 80.00%

Gambar 9. Hasil perhitungan akurasi keseluruhan data test.

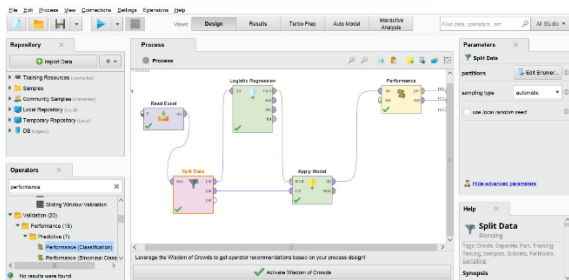
Akurasi keseluruhan model pada gambar 9 diatas menunjukkan bahwa hasil data uji adalah **80.00%**. Ini berarti bahwa model tersebut mampu memprediksi dengan benar 80% dari total sampel dalam data uji. Angka ini dihitung menggunakan fungsi `accuracy_score` dari `scikit-learn`, yang membandingkan hasil prediksi model (`model.predict(X_test)`) dengan label sebenarnya (`y_test`). Akurasi sebesar 80.00% menunjukkan performa yang cukup baik dari model dalam mengklasifikasikan data uji, meskipun masih ada ruang untuk perbaikan lebih lanjut jika diperlukan.



Gambar 10. Menampilkan visualisasi akurasi untuk setiap faktor.

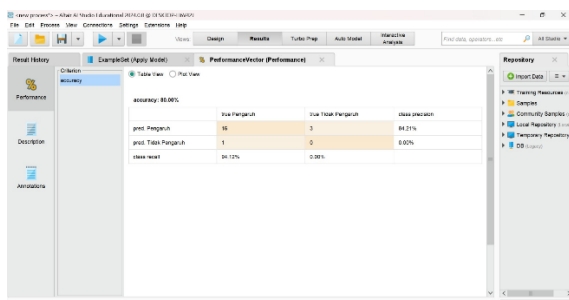
Berdasarkan visualisasi pada gambar 10 diatas, hasil akurasi pelatihan model regresi logistik untuk setiap faktor yang telah dianalisis sebelumnya. Setiap baris output menunjukkan nama faktor diikuti oleh persentase akurasi pelatihan model untuk faktor tersebut. Hasil yang didapat pada visualisasi diatas adalah Faktor Pendidikan memiliki persentase sebesar **86.25%**, Faktor Popularitas sebesar **86.25%** dan Faktor rekam jejak sebesar **87.50%**.

Pengujian Algoritma Regresi Logistik dengan RapidMiner



Gambar 11. Pemodelan Algoritma Regresi Logistik dengan RapidMiner.

Pemodelan seperti pada gambar 11 diatas menggunakan RapidMiner dengan menyertakan Operator Read Excel agar data dapat terbaca. Proses selanjutnya adalah menggunakan Operator split data, pada proses ini data akan dibagi menjadi dua bagian yaitu 80% data akan digunakan untuk menjadi data training, dan 20% data digunakan sebagai data testing. Selanjutnya, dimodelkan menggunakan Operator Algoritma Regresi Logistik dan dihubungkan dengan operator Apply model.



Gambar 12. Nilai Accuracy Algoritma Regresi Logistik dengan RapidMiner.

Gambar 12 diatas menunjukkan bahwa model memiliki akurasi sebesar 80% yang berarti 80% dari semua prediksi yang dibuat oleh model sesuai dengan nilai sebenarnya dari data uji. Akurasi ini menunjukkan seberapa baik model secara keseluruhan dalam klasifikasi.

KESIMPULAN

Berdasarkan analisis yang telah dilakukan, dapat disimpulkan bahwa :

1. Model klasifikasi yang diuji memiliki tingkat akurasi sebesar 85.00%. Model ini mengklasifikasikan dua kategori, yaitu "Pengaruh" dan "Tidak Pengaruh". Dari tabel yang disediakan, diketahui bahwa dari 17 sampel yang sebenarnya termasuk kategori "Pengaruh" (true Pengaruh), 15 sampel berhasil diklasifikasikan dengan benar sebagai "Pengaruh" (pred. Pengaruh), sementara 2 sampel salah diklasifikasikan sebagai "Tidak Pengaruh" (pred. Tidak Pengaruh). Untuk kategori "Tidak Pengaruh" (true Tidak Pengaruh), dari 3 sampel yang sebenarnya, hanya 2 yang diklasifikasikan dengan benar sebagai "Tidak Pengaruh" (pred. Tidak Pengaruh), dan 1 sampel salah diklasifikasikan sebagai "Pengaruh" (pred. Pengaruh). Precision untuk kategori "Pengaruh" adalah 93.75%, yang berarti dari semua sampel yang diklasifikasikan sebagai "Pengaruh" oleh model (16 sampel), 93.75% di antaranya benar-benar "Pengaruh". Precision untuk kategori "Tidak Pengaruh" adalah 50.00%, yang berarti dari semua sampel yang diklasifikasikan sebagai "Tidak Pengaruh" oleh model (4 sampel), 50.00% di antaranya benar-benar "Tidak Pengaruh". Recall untuk kategori "Pengaruh" adalah 88.24%, yang berarti dari semua sampel yang sebenarnya "Pengaruh" (17 sampel), 88.24% di antaranya diklasifikasikan dengan benar. Recall untuk kategori "Tidak Pengaruh" adalah 66.67%, yang berarti dari semua sampel yang sebenarnya "Tidak Pengaruh" (3 sampel), 66.67% di

- antaranya diklasifikasikan dengan benar. Secara keseluruhan, model ini lebih baik dalam mengidentifikasi sampel yang termasuk kategori "Pengaruh" dibandingkan dengan "Tidak Pengaruh".
2. Model regresi logistik yang dievaluasi pada data uji memperoleh tingkat akurasi sebesar 80.00%. Angka ini mencerminkan kemampuan model dalam memprediksi dengan benar 80% dari total sampel yang ada dalam dataset uji. Meskipun angka ini menunjukkan performa yang cukup baik dalam mengklasifikasikan data uji, ada potensi untuk melakukan peningkatan lebih lanjut melalui penyesuaian parameter atau evaluasi lebih mendalam terhadap fitur-fitur yang digunakan dalam model. Evaluasi terus-menerus diperlukan untuk memastikan bahwa model tetap dapat menghasilkan prediksi yang akurat.
 3. Hasil akurasi pelatihan model regresi logistik untuk setiap faktor yang telah dianalisis sebelumnya, dapat disimpulkan bahwa model menunjukkan kinerja yang baik dalam memprediksi data latih. Faktor Pendidikan memiliki akurasi sebesar **86.25%**, yang menunjukkan bahwa nilai tersebut memiliki persentase tinggi untuk memprediksi keputusan Pemilih. Faktor Popularitas juga menunjukkan akurasi sebesar **86.25%**, yang menyimpulkan bahwa popularitas memiliki urgensi yang hampir sama dengan pendidikan. Sementara itu, Faktor Rekam Jejak memiliki akurasi pelatihan tertinggi sebesar **87.50%**, dan memprediksi bahwa variabel yang terkait rekam jejak kandidat lebih efektif dalam membantu pemilih untuk memprediksi hasil yang benar dibandingkan dengan dua faktor

lainnya.

DAFTAR PUSTAKA

- [1] Tim Detikcom, "Pilgub DKI 2024 Tanggal Berapa? Ini Daftar Tahapannya," detik.com.
- [2] A. S. Fitriani, "JTAM (Jurnal Teori dan Aplikasi Matematika) Penerapan Data Mining Menggunakan Metode Klasifikasi Naïve Bayes untuk Memprediksi Partisipasi Pemilihan Gubernur," vol. 3, no. 2, pp. 98–104, 2019, doi: 10.31764/jtam.v3i2.995.
- [3] M. G. Sadewo, A. Perdana Windarto, I. S. Damanik, S. Tunas, and B. Pematangsiantar, "Prosiding Seminar Nasional Riset Information Science (SENARIS) Algoritma Naïve Bayes Dalam Memprediksi Kepuasan Nasabah," p. 318, 2019.
- [4] B. Purba and R. Syahputra, "Implementasi metode Naive Bayes Classifier pada Evaluasi Kepuasan Mahasiswa terhadap Pembelajaran Daring," vol. 6, no. 1, 2021, doi: 10.30743/infotekjar.v6i1.4352.
- [5] M. I. K. K. and D. S. Hendriyana, "Analisis Perbandingan Algoritma Support Vector Machine, Naive Bayes, Dan Regresi Logistik Untuk Memprediksi Donor Darah," *Jurnal Teknologi Terpadu*, vol. 8, no. 2, pp. 121–126, 2022.
- [6] Abd. C. Fauzan and K. Hikmah, "Implementasi Algoritma Naive Bayes Dalam Analisis Polarisasi Opini Masyarakat Terkait Vaksin Covid-19," *Rabit: Jurnal Teknologi dan Sistem Informasi Univrab*, vol. 7, no. 2, pp. 122–128, Jul. 2022, doi: 10.36341/rabit.v7i2.2403.
- [7] I. Muslim, K. Karo, M. Farhan, M. Fudzee, S. Kasim, and A. A. Ramli,

- “International Journal On Informatics Visualization Karonese Sentiment Analysis: A New Dataset and Preliminary Result,” 2022. [Online]. Available: www.joiv.org/index.php/joiv
- [8] I. M. K. Karo, M. F. M. Fudzee, S. Kasim, and A. A. Ramli, “Sentiment Analysis in Karonese Tweet using Machine Learning,” *Indonesian Journal of Electrical Engineering and Informatics*, vol. 10, no. 1, pp. 219–231, Mar. 2022, doi: 10.52549/ijeei.v10i1.3565.
- [9] C. N. Nasution and Y. Widyaningsih, “Klasifikasi Pemilih dalam Pemilu 2019 di Indonesia Menggunakan Regresi Logistik Multinomial dan Chi-Square Automatic Decision Tree (CHAID),” *Jurnal Statistika dan Aplikasinya*, vol. 6, no. 2, 2022.
- [10] R. Prabowo, H. Sujaini, and T. Rismawan, “Analisis Sentimen Pengguna Twitter Terhadap Kasus COVID-19 di Indonesia Menggunakan Metode Regresi Logistik Multinomial,” *Jurnal Sistem dan Teknologi Informasi (JustIN)*, vol. 11, no. 2, p. 366, Jul. 2023, doi: 10.26418/justin.v11i2.57449.
- [11] Syahril Dwi Prasetyo, Shofa Shofiah Hilabi, and Fitri Nurapriani, “Analisis Sentimen Relokasi Ibukota Nusantara Menggunakan Algoritma Naïve Bayes dan KNN,” *Jurnal KomtekInfo*, pp. 1–7, Jan. 2023, doi: 10.35134/komtekinfo.v10i1.330.
- [12] E. Novianto, A. Hermawan, and D. Avianto, “Klasifikasi Algoritma K-Nearest Neighbor, Naive Bayes, Decision Tree Untuk Prediksi Status Kelulusan Mahasiswa S1,” *Rabit : Jurnal Teknologi dan Sistem Informasi Univrab*, vol. 8, no. 2, pp. 146–154, Jul. 2023, doi: 10.36341/rabit.v8i2.3434.
- [13] F. D. Pratama, I. Zufria, and T. Triase, “Implementasi Data Mining Menggunakan Algoritma Naïve Bayes Untuk Klasifikasi Penerima Program Indonesia Pintar,” *Rabit : Jurnal Teknologi dan Sistem Informasi Univrab*, vol. 7, no. 1, pp. 77–84, Jan. 2022, doi: 10.36341/rabit.v7i1.2217.
- [14] M. Hasan, “Prediksi Tingkat Kelancaran Pembayaran Kredit Bank Menggunakan Algoritma Naive Bayes Berbasis Forward Selection,” *ILKOM Jurnal Ilmiah*, vol. 9, p. 317, 2017.
- [15] A. Pebdika, R. Herdiana, and D. Solihudin, “Klasifikasi Menggunakan Metode Naive Bayes Untuk Menentukan Calon Penerima PIP,” 2023.