

The Influence of Village Demographics on the Prediction of Community Participation in Elections Using the Naïve Bayes Algorithm

(Case Study: Pacitan City)

Pengaruh Demografi Desa Terhadap Prediksi Partisipasi Masyarakat Dalam Pemilu Menggunakan Algoritma Naïve Bayes (Studi Kasus : Kota Pacitan)

Ferry Setiawan¹, Arif Senja Fitriani², Ade Eviyanti³
{ ferry@gmail.com¹, asfjim@umsida.ac.id² }

Program Studi Informatika, Fakultas Sains dan Teknologi· Universitas Muhammadiyah Sidoarjo, Indonesia

Abstract- Voters' demographic factors have an influence in regional head elections or regional elections. The study will classify based on general election data obtained from villages in sub-districts in Pacitan Regency using data mining techniques. The variables that will be used in classifying villages are TPS, DPT, Attendance and Golput. The method that will be used is the Naïve Bayes method which is one of the classification techniques in data mining. Based on the research conducted, it is concluded that the information system created can classify villages into 2 types, namely low and high by Naïve Bayes method.

Keywords: Demographics, Classification, Naive Bayes.

Abstrak- Faktor demografi pemilih memiliki pengaruh dalam pemilihan umum kepala daerah atau pilkada. Penelitian akan melakukan klasifikasi berdasarkan data pemilihan umum yang diperoleh dari desa-desa di kecamatan yang ada di Kabupaten Pacitan dengan menggunakan teknik data mining. Variabel yang akan digunakan dalam melakukan klasifikasi desa adalah TPS, DPT, Kehadiran dan Golput. Metode yang akan digunakan adalah metode Naïve Bayes yang merupakan salah satu teknik pengklasifikasian dalam data mining. Berdasarkan penelitian yang dilakukan dihasilkan kesimpulan bahwa sistem informasi yang dibuat dapat mengklasifikasikan desa menjadi 2 jenis yaitu rendah dan tinggi menggunakan metode Naïve Bayes.

Kata kunci: Demografi, Klasifikasi, Naive Bayes

I. PENDAHULUAN

Demografi adalah istilah yang sering digunakan dalam pencatatan statistik. Arti demografi sendiri seringkali dikaitkan dengan kependudukan. Demografi adalah data statistik yang menyangkut populasi penduduk yang didasarkan atas berbagai klasifikasi seperti usia, ras, jenis kelamin, agama, pekerjaan, dan pendidikan. Lalu tingkat kelahiran, tingkat kematian, kepadatan penduduk, tingkat pendapatan, dan sebagainya. Demografi adalah data yang sangat penting dalam pengambilan kebijakan pemerintah. Pemerintah lazim menggunakan demografi untuk perencanaan kebijakan hingga pembagian sumber daya.[1]

Faktor demografi pemilih memiliki pengaruh dalam pemilu dan pilpres. Hanya saja, faktor demografi tersebut tidak cukup untuk memenangkan pilkada. Hak memilih bahkan dipandang sebagai jenis hak asasi manusia yang paling penting karena pelaksanaan berbagai jenis hak asasi manusia lainnya dapat dijamin melalui penggunaan hak pilih tersebut. Misalnya, pemilih hanya memberikan suara kepada partai/calon yang akan menjamin pelaksanaan berbagai jenis hak asasi manusia lainnya. Sebagaimana halnya setiap individu bertanggungjawab mengurus masalah pribadinya sendiri, demikian pula setiap warga negara sewajarnya berpartisipasi dalam pengurusan masalah publik dengan cara menggunakan hak pilihnya.[2]

II. METODE

Metode yang digunakan dalam penelitian ini terdiri dari 2, yaitu :

A. Data Mining dan Metode Naïve Bayes

Metode Naïve Bayes adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu class. Teorema Bayes yang memiliki kemampuan klasifikasi serupa dengan decision tree dan neural network.

Classification terbukti memiliki akurasi dan kecepatan yang tinggi saat diaplikasikan ke dalam database dengan data yang benar.[3]

B. Dasar Teorema Bayes

Dalam pengembangan Sistem Informasi "Klasifikasi Pengaruh Demografi Desa Terhadap Masyarakat Dalam Pemilu", terdapat beberapa konsep dasar untuk membangun Sistem Informasi ini[4], beberapa di antaranya yaitu:

C. Teorema Total Probabilitas

$$P(B) = \sum_{i=1}^M P(B|A_i)P(A_i) \dots\dots(1)$$

D. Teorema Bayes

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} = P(H) \times P(X|H) / P(X) \dots\dots(2)$$

- Misalkan X menjadi sampel data (bukti): label kelas tidak diketahui
- Misalkan H menjadi hipotesis bahwa X termasuk kelas C
- Klasifikasi adalah untuk menentukan $P(H|X)$, yakni probabilitas posteriori yang merupakan probabilitas yang dimiliki hipotesis yang diberikan dengan data sampel yang diamati X.
- $P(H)$ (probabilitas sebelumnya): probabilitas awal
Contoh: X akan membeli computer tanpa memandang usia, penghasilan
- $P(X)$: probabilitas bahwa data sampel diamati
- $P(X|H)$ (kemungkinan): probabilitas untuk mengamati sampel X, mengingat hipotesis itu berlaku.
Contoh: mengingat X akan membeli computer, masalahnya bahwa X adalah 31.. 40, penghasil menengah.

E. Prediksi Berdasarkan Teorema Bayes

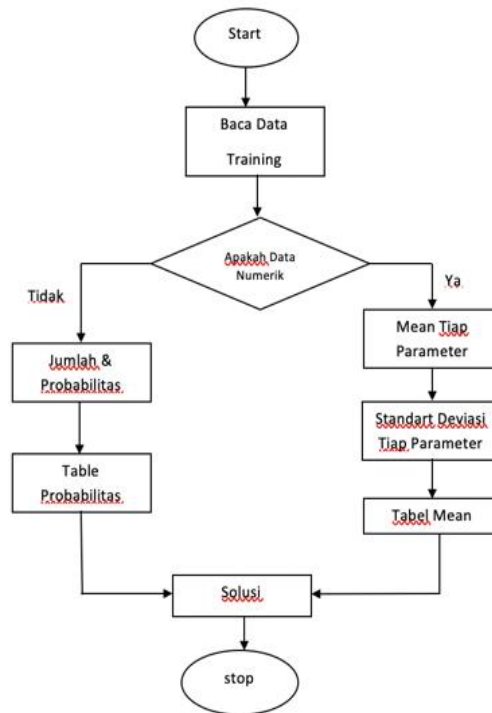
- Dengan data training X, probabilitas posterior dari hipotesis H, $P(H|X)$ mengikuti teorema Bayes

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} = P(H) \times P(X|H) / P(X) \dots\dots(3)$$

III. HASIL DAN PEMBAHASAN

Data yang digunakan dalam penelitian ini yakni data hasil pemilu 2021 kabupaten yang diperoleh dari website resmi KPU (<https://pemilu2019.kpu.go.id>).

Kabupaten Pacitan terdiri dari 12 kecamatan dan 166 desa (dari total 666 kecamatan dan 7.724 desa di Jawa Timur). jumlah penduduknya mencapai 582.275 jiwa dengan luas wilayah 1.389,92 km² dan sebaran penduduk 419 jiwa/km²[5]



Gambar 1. Flowchart Data Training

Perhitungan Manual Training Dataset

Tabel 1 Tabel Hasil Pemilu 2019 Kecamatan Pringkuku, Kabupaten Pacitan

ID	TPS	DPT	Hadir	Golput	Hadir %	Klasifikasi
1	Candi (4)	2627	1878	749	71.4883898	Tinggi
2	Dadapan (2)	1486	948	538	63.795424	Tinggi
3	Dersono (5)	2724	1611	1113	59.1409692	Tinggi
4	Glinggangan (3)	2166	1517	649	70.0369344	Tinggi
5	Jlubang (2)	1387	842	545	60.7065609	Tinggi
6	Ngadirejan (4)	3003	2033	970	67.6989677	Tinggi
7	Pelem (5)	3050	2139	911	70.1311475	Tinggi
8	Poko (3)	1757	1254	503	71.3716562	Tinggi
9	Pringkuku (4)	2775	1930	845	69.5495495	Tinggi
10	Sobo (1)	612	458	154	74.8366013	Tinggi
11	Sugiharas (2)	810	614	196	75.8024691	Tinggi
12	Tamanasri (3)	2059	1102	957	53.5211268	Rendah
13	Watukarung (2)	1222	930	292	76.1047463	Tinggi

Tabel 2 Tabel Klasifikasi Manual Sebagai Data Latih

ID	TPS	DPT	Hadir	Golput	Klasifikasi
1	< 4	High	Low	Low	Rendah
2	< 4	High	High	Low	Tinggi
3	4 - 7	High	Low	Low	Rendah
4	> 7	Medium	Low	Low	Rendah
5	> 7	Low	Low	High	Rendah
6	> 7	Low	High	High	Rendah
7	4 - 7	Low	High	High	Rendah
8	< 4	Medium	Low	Low	Rendah
9	< 4	Low	Low	High	Rendah
10	> 7	Medium	Low	High	Rendah
11	< 4	Medium	High	High	Rendah
12	4 - 7	Medium	High	Low	Tinggi
13	4 - 7	High	Low	High	Rendah
14	> 7	Medium	High	Low	Tinggi

- Class:
C1 = Rendah, C2 = Tinggi
- Data diklasifikasikan:
X = (TPS < 4, DPT = "Medium", Hadir = "Tinggi", Golput = "Tinggi")
- Apakah X → Desa dengan klasifikasi tinggi atau rendah?

Kalkulasi

- P(Ci):
 - P(klasifikasi desa = "Rendah") = 11/14 = 0.786
 - P(klasifikasi desa = "Tinggi") = 3/14 = 0.214
- Hitung P(X|Ci) untuk tiap kelas
 - P(TPS < 4 | klasifikasi desa = "Rendah") = 4/11 = 0.182
 - P(TPS < 4 | klasifikasi desa = "Tinggi") = 1/3 = 0.333
 - P(DPT = "Medium" | klasifikasi desa = "Rendah") = 4/11 = 0.182
 - P(DPT = "Medium" | klasifikasi desa = "Tinggi") = 2/3 = 0.667
 - P(Hadir = "High" | klasifikasi desa = "Rendah") = 3/11 = 0.273
 - P(Hadir = "High" | klasifikasi desa = "Tinggi") = 3/3 = 1
 - P(Golput = "High" | klasifikasi desa = "Rendah") = 7/11 = 0.636
 - P(Golput = "High" | klasifikasi desa = "Tinggi") = 0/3 = 0

X = (TPS < 4, DPT = "Medium", Hadir = "Tinggi", Golput = "Tinggi")

- P(X|Ci):
 - P(X | klasifikasi desa = "Rendah") = 0.182*0.182*0.273*0.636 = 0.006
 - P(X | klasifikasi desa = "Tinggi") = 0.333*0.667*1*0 = 0
- P(X|Ci) * P(Ci)
 - P(X | klasifikasi desa = "Rendah")*P(klasifikasi desa = "Rendah") = 0.006 * 0.786 = 0.005

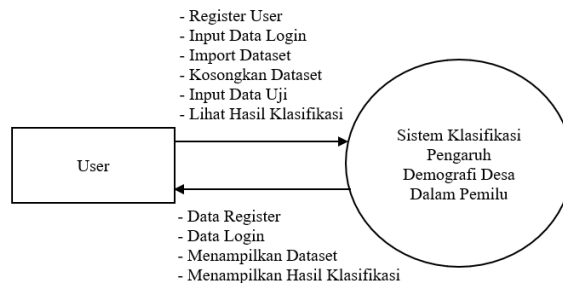
$$P(X | \text{klasifikasi desa} = \text{"Tinggi"}) * P(\text{klasifikasi desa} = \text{"Tinggi"}) \\ = 0 * 0.214 = 0$$

Oleh karena itu, X memiliki kelas (klasifikasi desa = "Rendah")

A. Diagram Konteks

Diagram konteks adalah diagram yang mencakup masukan-masukan dasar, sistem umum dan keluaran, diagram ini merupakan tingkatan tertinggi dalam diagram aliran data dan hanya memuat satu proses, menunjukkan sistem secara keseluruhan, diagram tersebut tidak memuat penyimpanan dan penggambaran aliran data yang sederhana, proses tersebut diberi nomor nol.[6]

Diagram konteks aplikasi ini adalah sebagai berikut:

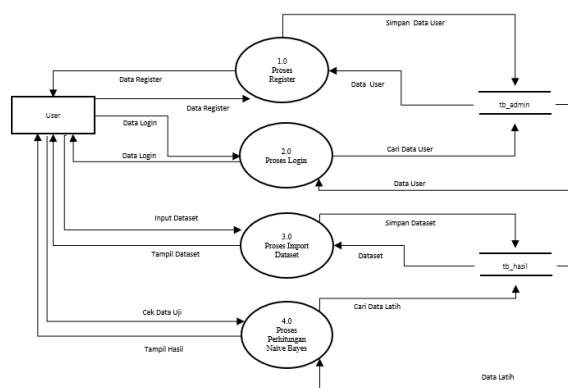


Gambar 2 Diagram Konteks

B. DFD Level 0

DFD menggambarkan penyimpanan data dan proses yang mentransformasikan data. DFD menunjukkan hubungan antara data dan proses pada sistem. DFD (Data Flow Diagram) adalah gambaran keseluruhan kerja sistem secara garis besar. DFD merupakan peralatan yang berfungsi untuk menggambarkan secara rinci mengenai sistem sebagai jaringan antar kerja dan kemana data mengalir serta penyimpanannya (Oetomo, 2002).

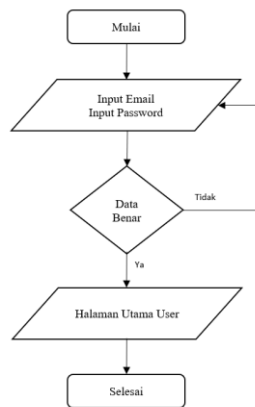
Data Flow Diagram (DFD) aplikasi ini sebagai berikut .



Gambar 3 Data Flow Diagram

C. Flowchart

Flowchart program digunakan untuk mengetahui bagaimana cara program dirancang. Flowchart dalam program akan terlihat jelas proses logika dalam suatu program dari aplikasi.

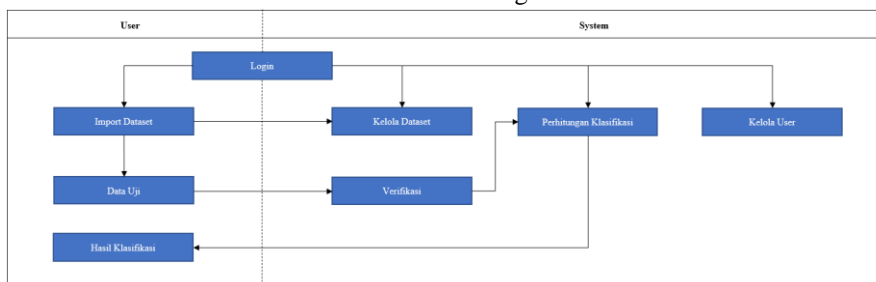


Gambar 4 Flowchart Login User

a. Struktur Navigasi

Struktur navigasi merupakan hal yang sangat penting dalam pembuatan sebuah aplikasi. Struktur navigasi yaitu rancangan hubungan antara beberapa form yang berbeda dan dapat membantu mengorganisasikan seluruh elemen input.[7]

Tabel 3 Struktur Navigasi



Gambar 3. 6 Struktur Navigasi Campuran

b. Perancangan Database

Pada perancangan database untuk Sistem Informasi ini terdiri dari 2 tabel yaitu table admin dan tabel hasil sebagai data latih.

- Entitas table admin (*primary key, auto increament: admin_id*)

Tabel 4 Tabel Data Latih Database Naïve Bayes

Nama field	Type data	Length	Keterangan
Admin_id	Integer	11	ID admin
Admin_email	Varchar	50	Email admin
Admin_password	Varchar	50	Password admin

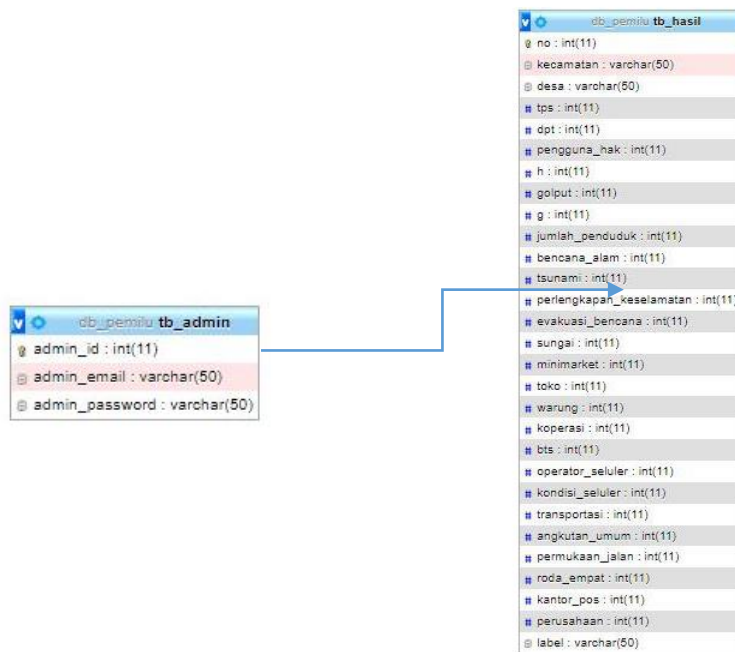
- Entitas data_latih (*primary key, auto increament: id*)

Tabel 5 Tabel Data Latih Database Naïve Bayes

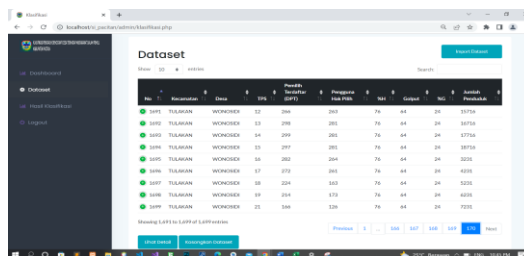
Nama field	Type data	Length	Keterangan
No	Integer	11	ID data latih
Kecamatan	Varchar	50	Jumlah kecamatan
Desa	Varchar	50	Jumlah desa
Tps	Integer	11	Jumlah tps
Dpt	Integer	11	Jumlah daftar pemilih tetap
Pengguna_hak	Integer	11	Jumlah DPT yang hadir

H	Integer	11	Jumlah presentase dpt
Golput	Integer	11	Jumlah golongan putih
G	Integer	11	Jumlah presentase golput
Jumlah_penduduk	Integer	11	Jumlah penduduk
Bencana_alam	Integer	11	Jumlah bencana alam
Tsunami	Integer	11	Jumlah tsunami
Perlengkapan_keselamatan	Integer	11	Jumlah perlengkapan
Evakuasi_bencana	Integer	11	Jumlah evakuasi bencana
Sungai	Integer	11	Jumlah sungai
Minimarket	Integer	11	Jumlah minimarket
Toko	Integer	11	Jumlah toko
Warung	Integer	11	Jumlah warung
Koperasi	Integer	11	Jumlah koperasi
Bts	Integer	11	Jumlah bts
Operator_seluler	Integer	11	Jumlah operator seluler
Kondisi_seluler	Integer	11	Jumlah kondisi seluler
Transportasi	Integer	11	Jumlah transportasi
Angkutan_umum	Integer	11	Jumlah angkutan umum
Permukaan_jalan	Integer	11	Jumlah permukaan jalan
Roda_empat	Integer	11	Jumlah roda empat
Kantor_pos	Integer	11	Jumlah kantor pos
Perusahaan	Integer	11	Jumlah perusahaan
Label	Varchar	50	Label klasifikasi desa

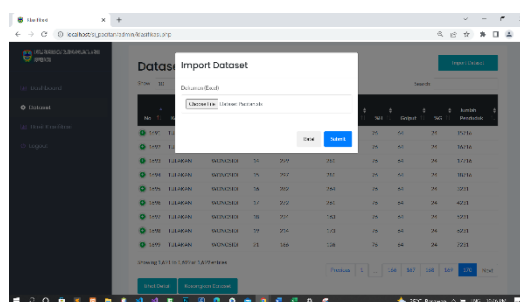
Tabel 6 Tabel Relasi Database Naïve Bayes



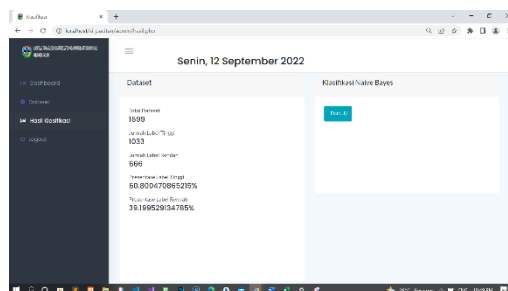
Berikut adalah tampilan aplikasi berbasis web :



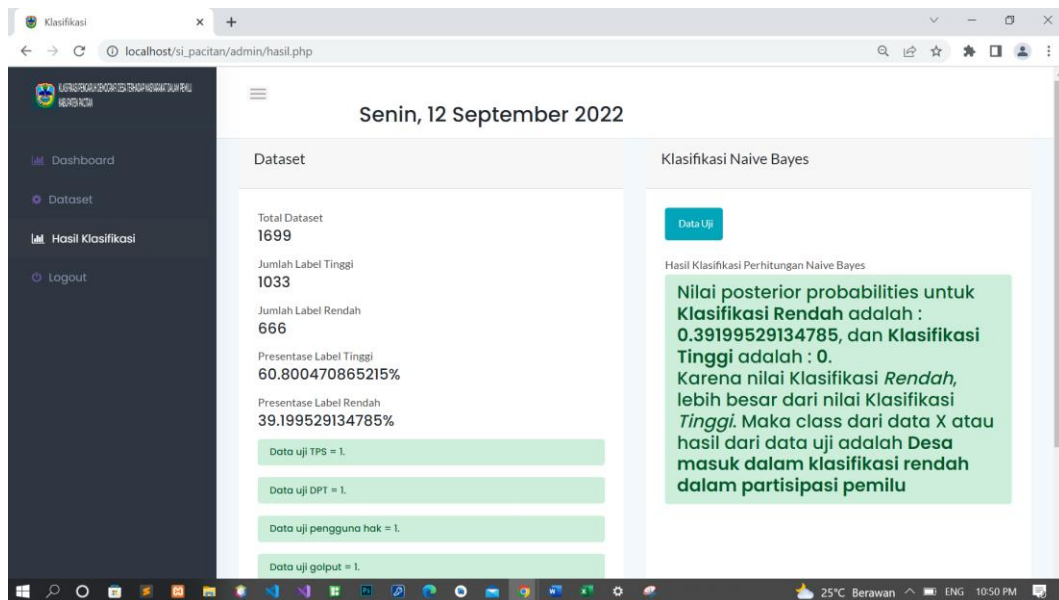
Gambar 5 Halaman Dataset



Gambar 6 Halaman Hasil Klasifikasi



Gambar 7 Form Input Data Uji



Gambar 8 Halaman Cek dan Hasil Data Uji

Dari Hasil dapat dijelaskan Bahwa tingkat klasifikasi tinggi menggunakan naïve bayes 60,8% dan tingkat klasifikasi rendah mencapai 39,1%.

IV. KESIMPULAN

Berdasarkan hasil permasalahan serta pembahasan yang telah diuraikan dalam bab sebelumnya, maka kesimpulan yang dapat di ambil adalah sebagai berikut:

- Telah dibuat dan di rancang Sistem Informasi Klasifikasi Pengaruh Demografi Desa Terhadap Masyarakat Dalam Pemilu menggunakan metode Naïve Bayes. Semoga sistem informasi ini dapat membantu pengguna untuk mempermudah dalam pengklasifikasian .
- Pada pengujian Black box testing dan pengujian sistem yang telah dibuat oleh peneliti menunjukkan penelitian ini telah sesuai dengan perancangan yang telah dikonsep sebelumnya yang terlihat dari hasil pengujian sehingga sistem informasi Klasifikasi Pengaruh Demografi Desa Terhadap Masyarakat Dalam Pemilu menggunakan metode Naïve Bayes ini siap untuk digunakan
- Hasil yang didapat dari penelitian ini dataset yang didapat 1699 data. Dengan jumlah mendapatkan label tinggi 1033 yang menghasilkan presentase lebel tinggi 60,8%. Sedangkan jumlah label rendah 666 menghasilkan presentase label rendah 39,1%

REFERENSI

- [1] Fitriani, A. S. (t.t.). Penerapan Data Mining Menggunakan Metode Klasifikasi Naïve Bayes untuk Memprediksi Partisipasi Pemilihan Gubernur.
- [2] Hakim, A. (2019). Prediksi Kehadiran Masyarakat Dalam Pemilihan Umum Dengan Menggunakan Metode Naïve Bayes Classification. 3.
- [3] Harahap, F., Saragih, N. E., Siregar, E. T., & Sariangah, H. (2021). PENERAPAN DATA MINING DENGAN ALGORITMA NAIVE BAYES CLASSIFIER DALAM MEMPREDIKSI PEMBELIAN CAT. JURNAL ILMIAH INFORMATIKA, 9(01), 19–23. <https://doi.org/10.33884/jif.v9i01.3702>
- [4] martha, layung pramesti. (2019). Hubungan karakteristik demografis masyarakat dengan tingkat partisipasi politik. <http://repository.unpak.ac.id>
- [5] Nurmayanti, W. P. (2021). Penerapan Naive Bayes dalam Mengklasifikasikan Masyarakat Miskin di Desa Lepak. Geodika: Jurnal Kajian Ilmu dan Pendidikan Geografi, 5(1), 123–132. <https://doi.org/10.29408/geodika.v5i1.3430>
- [6] Rifai, M. F., Jatnika, H., & Valentino, B. (2019). Penerapan Algoritma Naïve Bayes Pada Sistem Prediksi Tingkat Kelulusan Peserta Sertifikasi Microsoft Office Specialist (MOS). PETIR, 12(2), 131–144. <https://doi.org/10.33322/petir.v12i2.471>
- [7] Sabrani, A., & Majapahit, J. (2020). METODE MULTINOMIAL NAÏVE BAYES UNTUK KLASIFIKASI ARTIKEL ONLINE TENTANG GEMPA DI INDONESIA. 2(1).

- [8] Silitonga, W. H., & Sihotang, J. I. (2019). Analisis Sentimen Pemilihan Presiden Indonesia Tahun 2019 Di Twitter Berdasarkan Geolocation Menggunakan Metode Naïve Bayesian Classification. *TelKa*, 9(02), 115–127. <https://doi.org/10.36342/teika.v9i02.2199>
- [9] Wahyuni, D. T., Sutojo, T., & Luthfiarta, A. (t.t.). PREDIKSI HASIL PEMILU LEGISLATIF DKI JAKARTA MENGGUNAKAN NAÏVE BAYES DENGAN ALGORITMA GENETIKA SEBAGAI FITUR SELEKSI.
- [10] Wijaya, H. D., & Dwiasnati, S. (2020). Implementasi Data Mining dengan Algoritma Naïve Bayes pada Penjualan Obat. *Jurnal Informatika*, 7(1), 1–7. <https://doi.org/10.31311/ji.v7i1.6203>