

The Influence of Village Demographics on the Prediction of Community Participation in Elections Using the Naive Bayes Algorithm (Case Study: Pacitan City)

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

Ferry Setiawan¹⁾, Arif Senja Fitriani²⁾

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

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

Email Penulis Korespondensi: asfjim@umsida.ac.id²⁾

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 yaitu Agama, suku, daerah dan imigrasi dikarenakan 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]

Penelitian terdahulu menjadi salah satu referensi penelitian penulis, sehingga penulis dapat memperkaya teori yang digunakan untuk mengkaji penelitian yang dilakukan. Dari penelitian terdahulu, penulis tidak menemukan penelitian dengan judul yang sama seperti penelitian penulis, akan tetapi penulis telah mengajukan beberapa penelitian sebagai referensi untuk memperkaya bahan dalam penelitian penulis.

Penelitian yang dilakukan oleh Ailul Chowiyah (2019) yang berjudul "PENERAPAN DATA MINING MENGGUNAKAN METODE KLASIFIKASI NAÏVE BAYES UNTUK MEMPREDIKSI PARTISIPASI PEMILIHAN GUBERNUR". Tujuan dari penelitian tersebut yaitu untuk mengetahui nilai klasifikasi algoritma Naïve Bayes dalam klasifikasi memprediksi partisipasi pemilihan Gubernur. Hasil yang didapatkan dalam penelitian ini yang digunakan yaitu dataset yang diambil sebanyak 300 data dibagi 2 yaitu sebanyak 65% dari 195 data training dan sebanyak 35% dari 105 data Testing. Hasil prediksi partisipasi pemilu dari dataset yang diambil sebanyak 300 data dibagi 2 yaitu sebanyak 65% dari 195 data training dan sebanyak 35% dari 105 data testing. Hasil prediksi berdasarkan set atribut kehadiran, untuk kehadiran dengan 105 data diperoleh nilai 97% prediksi kebenarannya dan diperoleh 3% prediksi kesalahannya.

Penelitian lainnya dilakukan oleh Amelia Iomenta & Jacobus (2017) yang laporan berjudul “PREDIKSI MASA STUDI MAHASISWA MENGGUNAKAN ALGORITMA NAÏVE BAYES”. Tujuan dari penelitian penulis yaitu untuk menganalisa dan memprediksi lama masa studi mahasiswa sehingga dapat dijadikan acuan dalam pengadaan sarana dan prasarana perkuliahan. Hasil yang didapat dari penelitian ini setelah dilakukan pengujian, diperoleh nilai akurasi sebesar 82,26%. Hasil tersebut dapat dimanfaatkan sebagai strategi dalam meningkatkan kualitas pembelajaran di Universitas Mercu Buana.

Penelitian yang dilakukan oleh Abdul Hakim tahun (2019) yang berjudul “PREDIKSI KEHADIRAN MASYARAKAT DALAM PEMILIHAN UMUM DENGAN MENGGUNAKAN METODE NAÏVE BAYES”. Tujuan dari penelitian tersebut yaitu untuk mengetahui nilai klasifikasi algoritma Naïve Bayes dalam klasifikasi dalam kehadiran masyarakat dalam pemilu. Serta hasil yang didapatkan dalam penelitian ini nilai accuracy sebesar 68,17%, nilai untuk precision sebesar 77,46%, dan nilai untuk recall sebesar 78,57%. Dengan data training 80% dan data test 20%. Penelitian ini memiliki kesamaan yakni sama menggunakan metode Naïve Bayes selain itu terdapat perbedaan dalam dataset. Pada penelitian yang telah dilakukan oleh Abdul Hakim (2019) menggunakan algoritma Naives Bayes.

II. METODE

Metode yang digunakan dalam penelitian ini, yaitu :

A. 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. Teorema Bayes

Bayes merupakan suatu teknik prediksi yang berbasis peluang (probabilistik) sangat sederhana yang berdasarkan pada penerapan aturan teori Bayes dengan asumsi ketidakbergantungan (independensi) yang kuat (naif). Dalam pengembangan Sistem Informasi, terdapat beberapa konsep dasar untuk membangun Sistem Informasi ini[4], beberapa di antaranya yaitu:

1. Teorema Total Probabilitas

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

2. Teorema Bayes

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

- 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.

3. Prediksi Berdasarkan Teorema Bayes

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

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

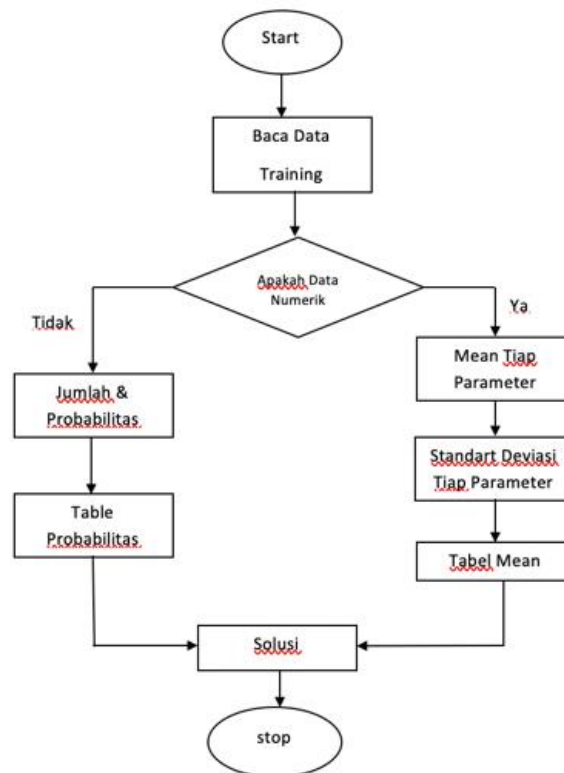
III. HASIL DAN PEMBAHASAN

Data yang digunakan dalam penelitian ini yakni data hasil pemilu 2019 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]

A. Data Training

Sumber data training adalah data dari hasil pilpres 2019 di kabupaten pacitan. Jumlah data hasil pilpres 2019 yang keseluruhannya adalah 1699 data.



Gambar 1. Flowchart Data Training

Tabel 2 Tabel Klasifikasi Manual

| 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:
 $X = (TPS < 4, DPT = \text{"Medium"}, \text{Hadir} = \text{"Tinggi"}, \text{Golput} = \text{"Tinggi"})$
- Apakah $X \rightarrow$ Desa dengan klasifikasi tinggi atau rendah?
Kalkulasi
- $P(C_i)$:
 - $P(\text{klasifikasi desa} = \text{"Rendah"}) = 11/14 = 0.786$
 - $P(\text{klasifikasi desa} = \text{"Tinggi"}) = 3/14 = 0.214$
- Hitung $P(X|C_i)$ untuk tiap kelas
 - $P(TPS < 4 | \text{klasifikasi desa} = \text{"Rendah"}) = 4/11 = 0.182$
 - $P(TPS < 4 | \text{klasifikasi desa} = \text{"Tinggi"}) = 1/3 = 0.333$
 - $P(DPT = \text{"Medium"} | \text{klasifikasi desa} = \text{"Rendah"}) = 4/11 = 0.182$
 - $P(DPT = \text{"Medium"} | \text{klasifikasi desa} = \text{"Tinggi"}) = 2/3 = 0.667$
 - $P(\text{Hadir} = \text{"High"} | \text{klasifikasi desa} = \text{"Rendah"}) = 3/11 = 0.273$
 - $P(\text{Hadir} = \text{"High"} | \text{klasifikasi desa} = \text{"Tinggi"}) = 3/3 = 1$
 - $P(\text{Golput} = \text{"High"} | \text{klasifikasi desa} = \text{"Rendah"}) = 7/11 = 0.636$
 - $P(\text{Golput} = \text{"High"} | \text{klasifikasi desa} = \text{"Tinggi"}) = 0/3 = 0$

$X = (TPS < 4, DPT = \text{"Medium"}, \text{Hadir} = \text{"Tinggi"}, \text{Golput} = \text{"Tinggi"})$

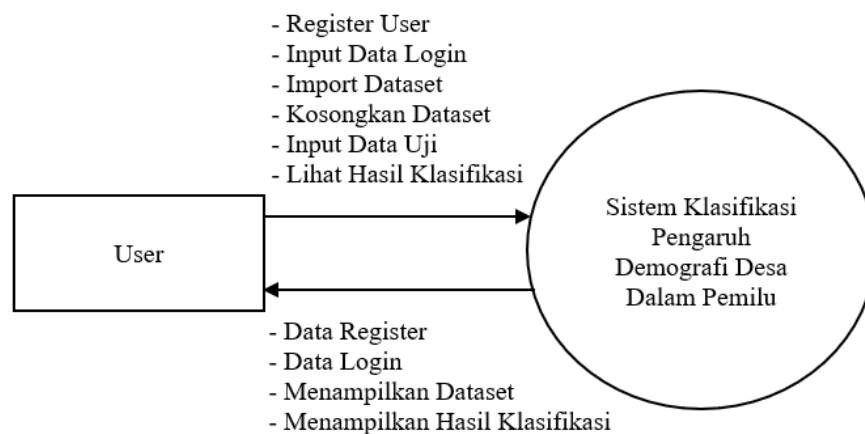
- $P(X|C_i)$:
 - $P(X | \text{klasifikasi desa} = \text{"Rendah"}) = 0.182 * 0.182 * 0.273 * 0.636 = 0.006$
 - $P(X | \text{klasifikasi desa} = \text{"Tinggi"}) = 0.333 * 0.667 * 1 * 0 = 0$
- $P(X|C_i) * P(C_i)$
 - $P(X | \text{klasifikasi desa} = \text{"Rendah"}) * P(\text{klasifikasi desa} = \text{"Rendah"}) = 0.006 * 0.786 = 0.005$
 - $P(X | \text{klasifikasi desa} = \text{"Tinggi"}) * P(\text{klasifikasi desa} = \text{"Tinggi"}) = 0 * 0.214 = 0$

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:

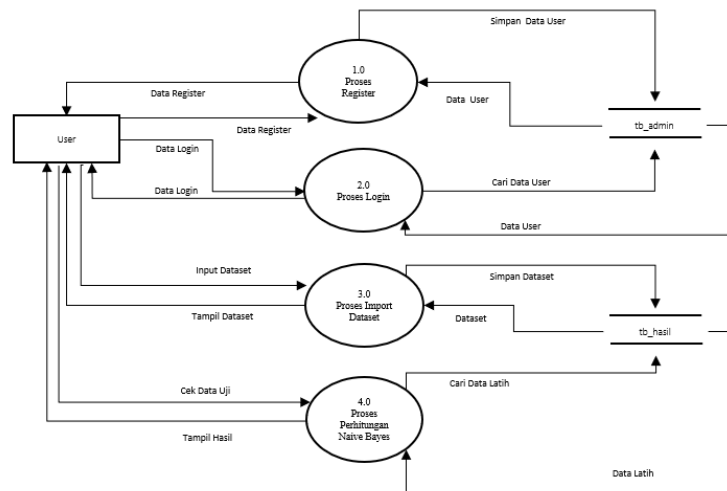


Gambar 2 Diagram Konteks

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).

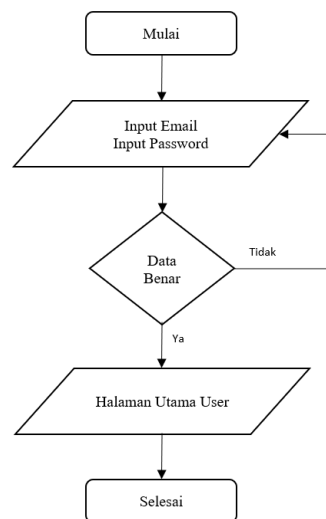
Data Flow Diagram (DFD) aplikasi ini sebagai berikut .



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.

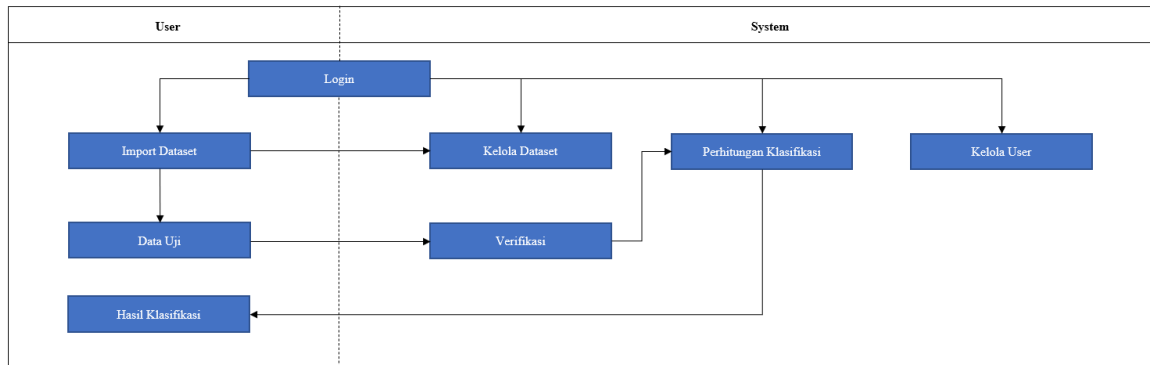


Gambar 4 Flowchart Login User

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

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 increment: 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 |

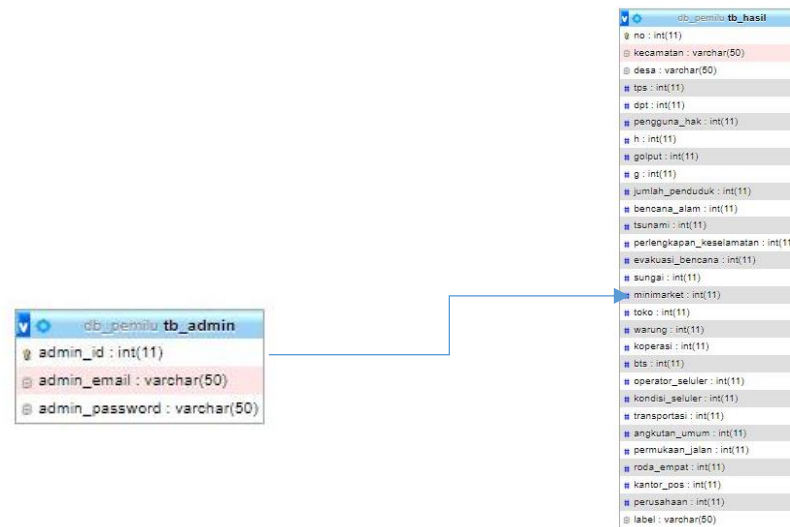
- Entitas data_latih (*primary key, auto increment: id*)

Tabel 5 Tabel Data Latih Database Naïve Bayes

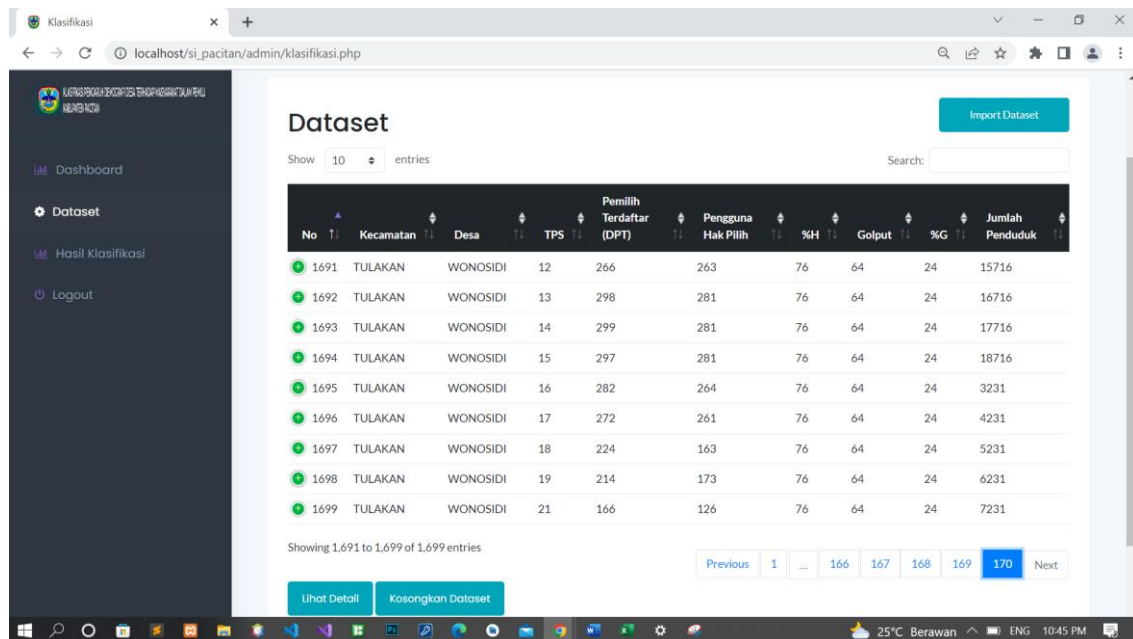
| Nama field | Type data | Length | Keterangan |
|-----------------|-----------|--------|-----------------------------|
| No | Integer | 11 | ID data latih |
| Kecamatan | Varchar | 50 | Jumlah kecamatan |
| Desa | Varchar | 50 | Jumlah desa |
| Tps | Integer | 11 | Jumlah tps |
| Dpt | Integer | 11 | 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 |

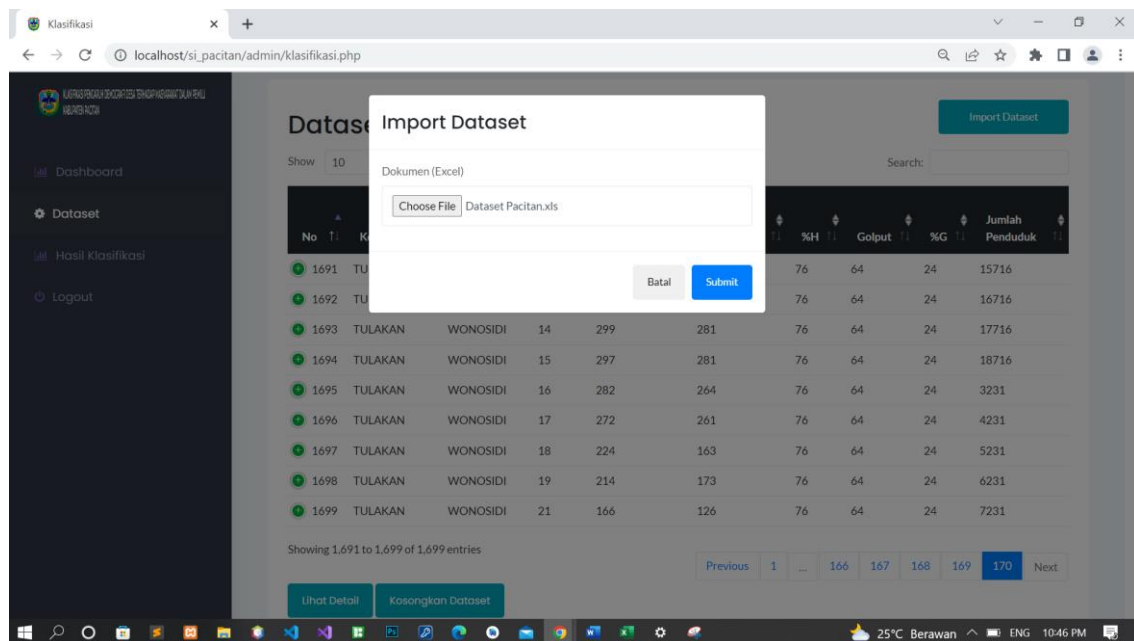
Tabel 6 Tabel Relasi Database Naïve Bayes



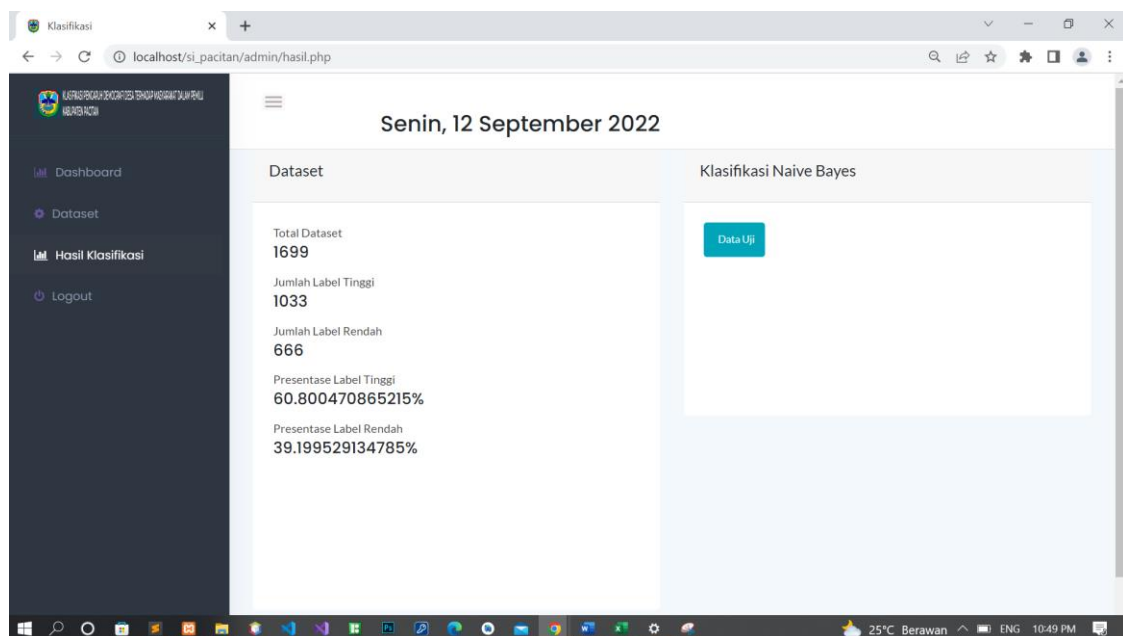
Berikut adalah tampilan aplikasi berbasis web :



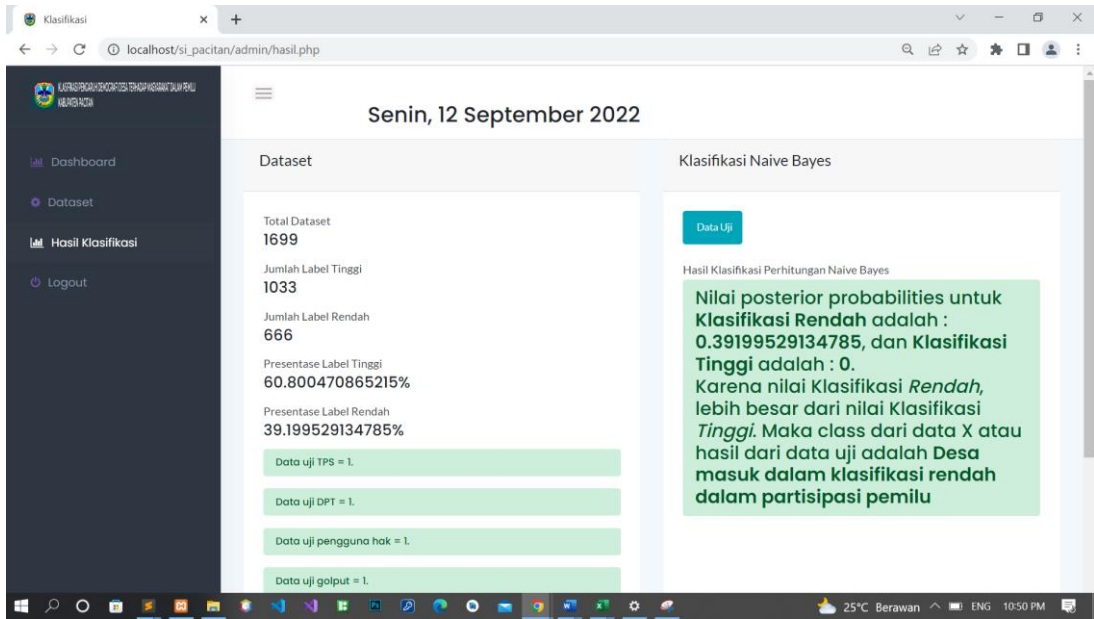
Gambar 5 Halaman Dataset



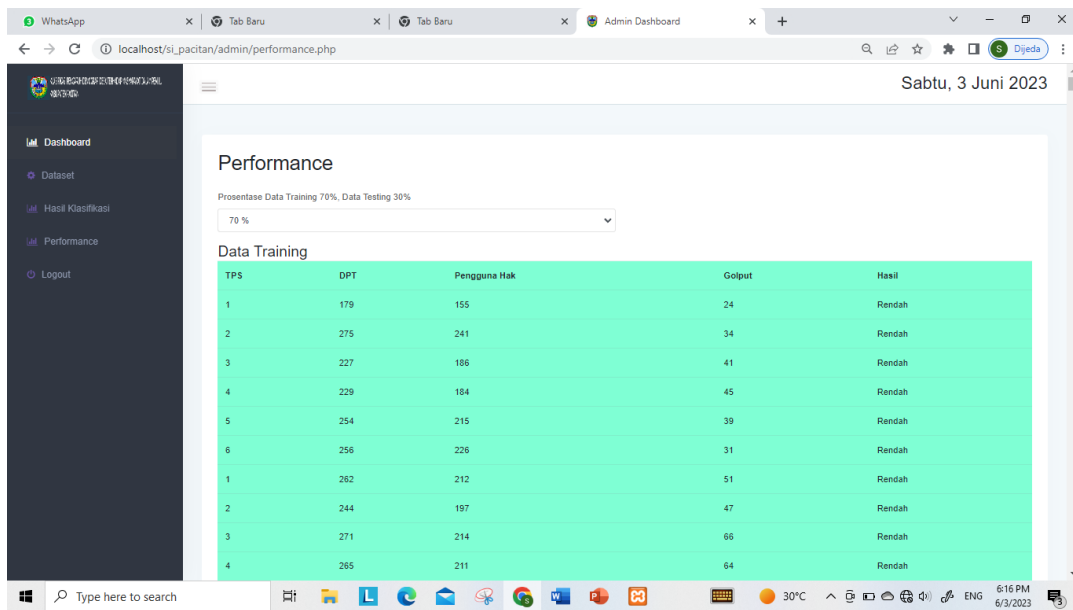
Gambar 6 Halaman Hasil Klasifikasi



Gambar 7 Form Input Data Uji



Gambar 8 Halaman Cek dan Hasil Data Uji Klasifikasi



Gambar 9 Halaman performance

| ID | Value 1 | Value 2 | Value 3 | Value 4 | Value 5 |
|----|---------|---------|---------|---------|---------|
| 8 | 299 | 285 | 64 | Tinggi | Tinggi |
| 9 | 283 | 278 | 64 | Tinggi | Tinggi |
| 11 | 284 | 277 | 64 | Tinggi | Tinggi |
| 11 | 261 | 255 | 64 | Tinggi | Tinggi |
| 12 | 266 | 263 | 64 | Tinggi | Tinggi |
| 13 | 298 | 281 | 64 | Tinggi | Tinggi |
| 14 | 299 | 281 | 64 | Tinggi | Tinggi |
| 15 | 297 | 281 | 64 | Tinggi | Tinggi |
| 16 | 282 | 264 | 64 | Tinggi | Tinggi |
| 17 | 272 | 261 | 64 | Tinggi | Tinggi |
| 18 | 224 | 163 | 64 | Rendah | Tinggi |
| 19 | 214 | 173 | 64 | Rendah | Tinggi |
| 21 | 166 | 126 | 64 | Rendah | Tinggi |

Akurasi : 64.31%

Gambar 9 cek hasil data uji data training dan data testing

Dari Hasil dapat dijelaskan Bahwa hasil dari pengambilan data training 70% dan data testing 30% dari 1699 data set yang diambil mendapat akurasi 64,31%

Tabel 7 Berdasarkan perhitungan web

| Prediksi | Data Training | Data Testing | Akurasi Presentase % |
|----------|---------------|--------------|----------------------|
| Web | 70% | 30% | 64,31% |

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 .
- Hasil yang didapat dari penelitian. Dari dataset yang didapat 1699 data dan akurasi yang didapat dengan menguji data training 70% dan data testing 30% mencapai akurasi 64,31%

V. Ucapan Terimakasih

Ucapan terimakasih kepada bapak dosen pembimbing, Mentor, Serta perguruan tinggi Universitas Muhammadiyah Sidoarjo yang telah membantu dan mendukung dalam penyelesaian penulisan artikel ini dan mendukung dalam menyelesaikan penulisan artikel ini dan terimakasih kepada pihak editor yang telah mereview dan menelaah penulisan artikel ini. Besar harapan kami untuk mendapat respon yang baik dari jurnal ilmiah Teknik informatika dalam publikasi jurnal tersebut.

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://repsitory.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>

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.