

BAB IV HASIL DAN PEMBAHASAN

4.1. Deskripsi Data

Data yang digunakan dalam penelitian merupakan data sekunder yang diperoleh dari Kantor Desa Tembung, Kecamatan Percut Sei Tuan. Data yang digunakan berupa data rumah tangga di Desa Tembung tahun 2023 yang memiliki 2 kelas yaitu miskin dan non miskin.

Jumlah populasi dalam penelitian ini ada sebanyak 13.099 data. Teknik pengambilan sampel menggunakan *purposive sampling* (Isaac, 2023). Lalu penentuan jumlah sampel ditentukan menggunakan rumus slovin (Wardhani & Kamil, 2023). Berikut perhitungan menggunakan rumus slovin dengan margin error toleransi 5%:

$$n = \frac{N}{1 + Ne^2}$$

$$n = \frac{13099}{1 + 13099(0.05)^2}$$

$$n = 388.147 \text{ dibulatkan menjadi } n = 390$$

di mana:

n = jumlah sampel

N = jumlah populasi

e = margin error toleransi

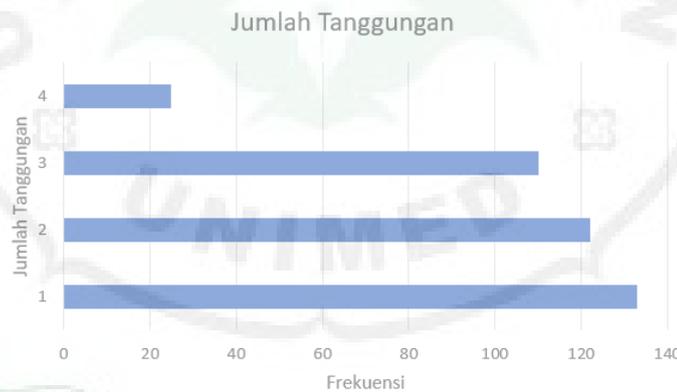
Sehingga dari perhitungan menggunakan rumus slovin tersebut didapatkan jumlah sampel yang optimal yaitu sebanyak 390 data. Data yang digunakan merupakan data rumah tangga yang meliputi jumlah tanggungan (A_1), pendidikan terakhir (A_2), pekerjaan (A_3), penghasilan (A_4), jumlah mobil (A_5), jumlah sepeda motor (A_6), status kepemilikan rumah (A_7), kondisi rumah (A_8) dan label (B). Berikut ini adalah data sampel yang digunakan pada penelitian ini.

Tabel 4. 1. Sampel Data Penduduk Desa Tembung

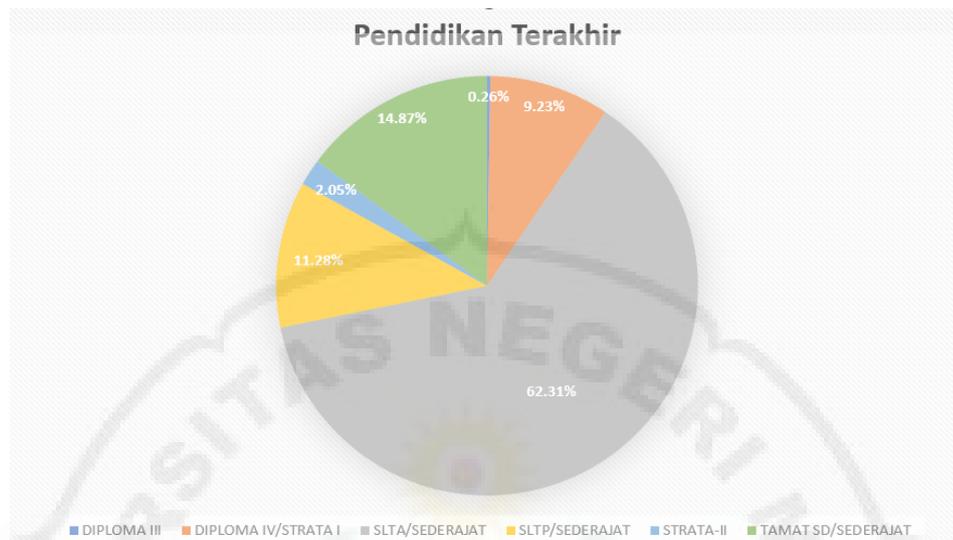
No.	A1	A2	...	A5	A6	A7	A8	B
1	1	Tamat SD/Sederajat	...	0	0	Milik Sendiri	Semi Permanen	Miskin
2	4	Tamat SD/Sederajat	...	0	1	Milik Sendiri	Semi Permanen	Miskin
3	2	Tamat SD/Sederajat	...	0	1	Milik Sendiri	Semi Permanen	Miskin
...
388	3	SLTA/Sederajat	...	0	1	Milik Sendiri	Semi Permanen	Non-Miskin
389	2	Diploma-IV/Strata-I	...	0	2	Milik Sendiri	Permanen	Non-Miskin
390	2	SLTA/Sederajat	...	0	1	Milik Sendiri	Permanen	Non-Miskin

(Sumber Data: Kantor Kepala Desa Tembung, 2023)

Tabel 4.1 menampilkan sampel data penduduk Desa Tembung yang diperoleh dari Kantor Kepala Desa Tembung. Kemudian berikut ini merupakan deskripsi data per variabel yang digunakan pada penelitian ini.

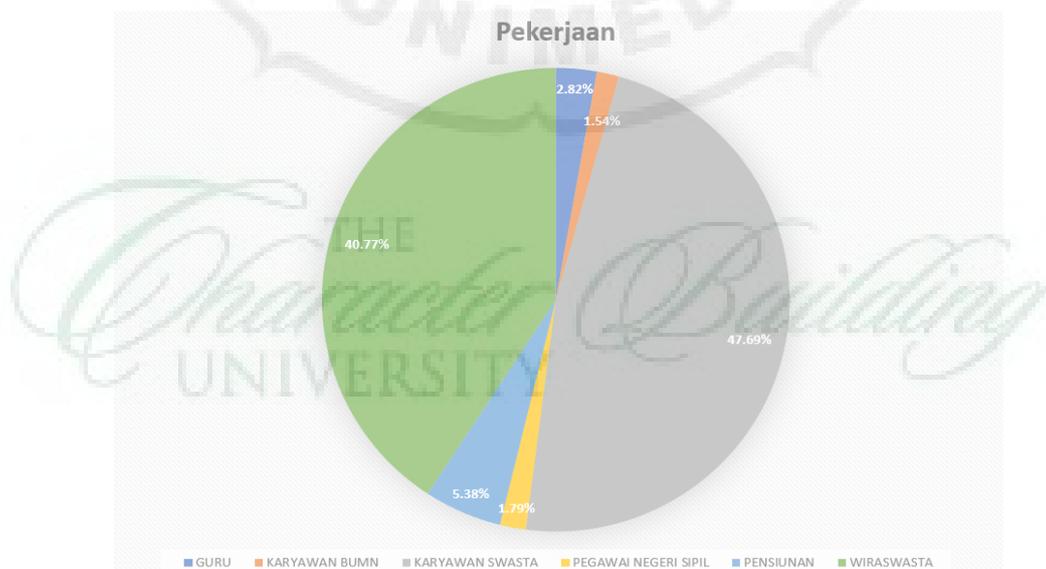
**Gambar 4. 1.** Diagram Variabel Jumlah Tanggungan

Gambar 4.1 menampilkan diagram variabel jumlah tanggungan dimana sumbu y merupakan jumlah tanggungan yang muncul dan sumbu x adalah frekuensi kemunculan. Dari gambar 4.1 terlihat bahwa variabel jumlah tanggungan (A_1) memiliki empat kategori yang muncul, yaitu $A_1 = 1$, $A_1 = 2$, $A_1 = 3$ dan $A_1 = 4$. Kategori $A_1 = 1$ merupakan kategori dengan frekuensi kemunculan yang paling banyak yaitu sebanyak 133 data. Kemudian kategori $A_1 = 4$ merupakan kategori dengan frekuensi kemunculan yang paling sedikit yaitu sebanyak 25 data.



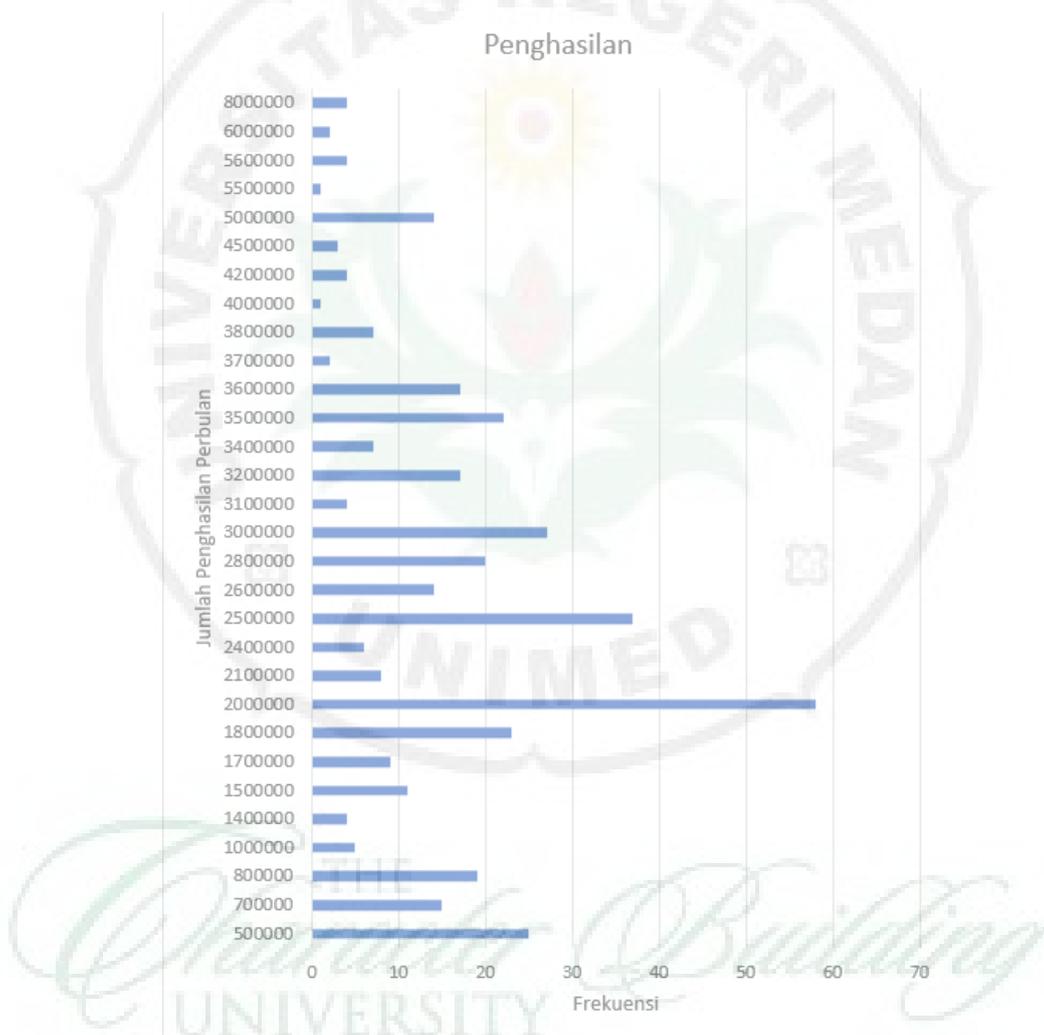
Gambar 4. 2. Diagram Variabel Pendidikan Terakhir

Dari gambar 4.2 terlihat bahwa variabel pendidikan terakhir (A_2) memiliki enam kategori yang muncul dimana setiap warna mewakili setiap kategori yang ada dalam variabel pendidikan terakhir. Kategori SLTA/Sederajat merupakan kategori dengan frekuensi kemunculan yang paling banyak yaitu sebanyak 62.31% (243 data). Kemudian kategori Diploma-III merupakan kategori dengan frekuensi kemunculan yang paling sedikit yaitu sebanyak 0.26% (1 data).



Gambar 4. 3. Diagram Variabel Pekerjaan

Dari gambar 4.3 terlihat bahwa variabel pekerjaan (A_3) memiliki enam kategori yang muncul dimana setiap warna mewakili setiap kategori yang ada dalam variabel pekerjaan. Kategori karyawan swasta merupakan kategori yang memiliki frekuensi kemunculan yang paling banyak yaitu sebanyak 47.69% (189 data). Kemudian kategori karyawan BUMN merupakan kategori dengan frekuensi kemunculan yang paling sedikit yaitu sebanyak 1.54% (6 data).



Gambar 4. 4. Diagram Variabel Penghasilan

Gambar 4.4 menampilkan diagram variabel penghasilan dimana sumbu x adalah frekuensi kemunculan dan sumbu y merupakan besar penghasilan perbulan. Dari gambar 4.4 terlihat bahwa variabel penghasilan (A_4) memiliki tiga puluh kategori yang muncul. Kategori $A_4 = 2000000$ merupakan kategori dengan frekuensi kemunculan yang paling banyak yaitu sebanyak 58 data. Kemudian

kategori $A_4 = 5500000$ dan kategori $A_4 = 4000000$ merupakan kategori dengan frekuensi kemunculan yang paling sedikit yaitu sebanyak 1 data.



Gambar 4. 5. Diagram Variabel Jumlah Mobil

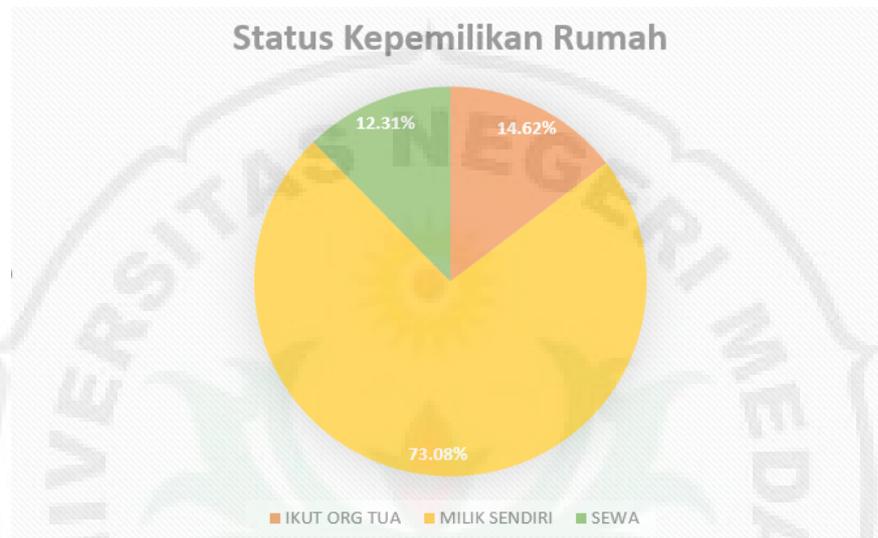
Gambar 4.5 menampilkan diagram variabel jumlah mobil sumbu x adalah frekuensi kemunculan dan sumbu y merupakan jumlah mobil yang dimiliki. Dari gambar 4.5 terlihat bahwa variabel jumlah mobil (A_5) hanya memiliki dua kategori yang muncul, yaitu $A_5 = 0$ dan $A_5 = 1$ dengan masing-masing frekuensi kemunculannya yaitu sebanyak 357 data untuk kategori $A_5 = 0$ dan sebanyak 33 data untuk kategori $A_5 = 1$.



Gambar 4. 6. Diagram Variabel Jumlah Motor

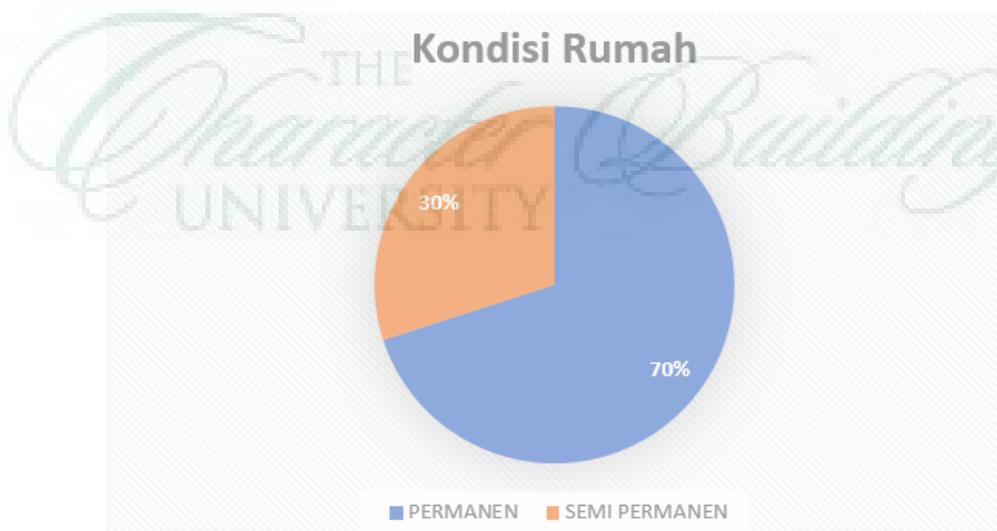
Gambar 4.6 menampilkan diagram variabel jumlah motor dimana sumbu x adalah frekuensi kemunculan dan sumbu y merupakan jumlah sepeda motor yang dimiliki. Dari gambar 4.6 terlihat bahwa variabel jumlah motor (A_6) memiliki empat kategori yang muncul. Kategori $A_6 = 1$ merupakan kategori dengan frekuensi

kemunculan yang paling banyak yaitu sebanyak 208 data. Kemudian kategori $A_6 = 3$ merupakan kategori dengan frekuensi kemunculan yang paling sedikit yaitu sebanyak 9 data.



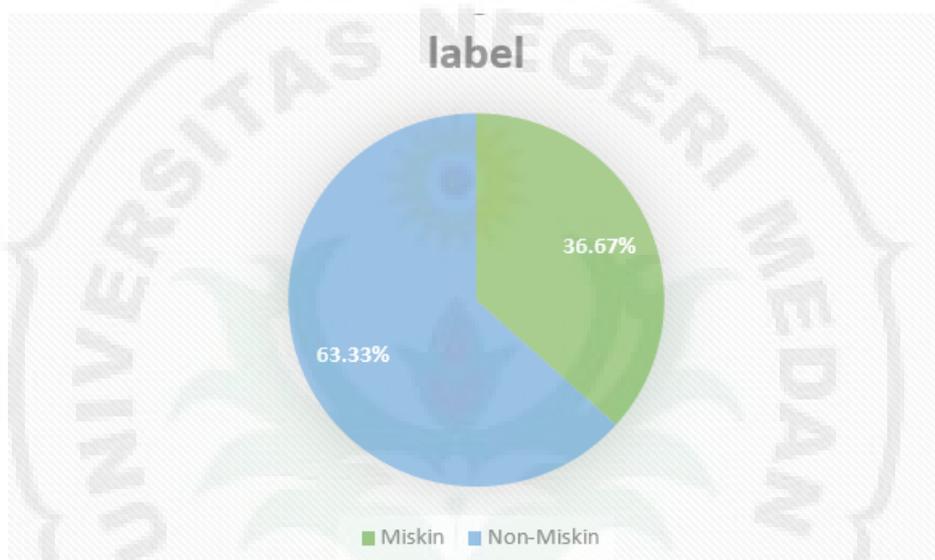
Gambar 4. 7. Diagram Variabel Status Kepemilikan Rumah

Dari gambar 4.7 terlihat bahwa variabel status kepemilikan rumah (A_7) memiliki tiga kategori yang muncul dimana setiap warna mewakili setiap kategori yang ada dalam variabel status kepemilikan rumah. Kategori milik sendiri merupakan kategori dengan frekuensi kemunculan yang paling banyak yaitu sebanyak 74.08% (285 data). Kemudian kategori sewa merupakan kategori dengan frekuensi kemunculan yang paling sedikit yaitu sebanyak 12.31% (48 data).



Gambar 4. 8. Diagram Variabel Kondisi Rumah

Dari gambar 4.8 terlihat bahwa variabel kondisi rumah (A_8) memiliki dua kategori yang muncul dimana setiap warna mewakili setiap kategori yang ada dalam variabel kondisi rumah, yaitu kategori semi permanen dan kategori permanen dengan masing-masing frekuensi kemunculan sebanyak 30% (117 data) untuk kategori semi permanen dan sebanyak 70% (273 data) untuk kategori permanen.



Gambar 4. 9. Diagram Variabel Label

Dari gambar 4.8 terlihat bahwa variabel label (B) hanya memiliki dua kelas, yaitu kelas miskin dan kelas non-miskin dengan masing-masing frekuensi kemunculan sebanyak 36.67% (143 data) untuk kelas miskin dan sebanyak 63.33% (247 data) untuk kelas non-miskin.

4.2. Ukuran *Split Data*

Setelah data dikumpulkan, kemudian sampel dibagi menjadi data *train* dan data *test* dengan 3 perbandingan yang paling sering digunakan, yaitu 70:30, 80:20 dan 90:10, untuk mengetahui ukuran perbandingan yang paling cocok akan dinilai dengan melihat score dari *k-fold cross validation* dimana nilai k yang digunakan yaitu $k = 5$, $k = 10$, $k = 15$, $k = 20$.

K-Fold Cross Validation adalah salah satu jenis pengujian *cross-validation* yang berfungsi untuk menilai kinerja proses algoritma dengan membagi kumpulan data sampel secara acak dan mengelompokkan data sebanyak nilai K pada *K-Fold*. Dalam metode *K-Fold Cross Validation*, kumpulan data yang telah dibagi secara

acak diolah dalam beberapa percobaan K , di mana data partisi K digunakan sebagai data uji dan sisa bagian digunakan sebagai data pelatihan (Kurniawan, 2017). Kemudian model terbaik dipilih setelah membandingkan semua akurasi yang dihasilkan yang dapat dilihat pada tabel 4.1.

Tabel 4. 2. Hasil akurasi untuk setiap ukuran *split* data

Ukuran <i>split</i> data	Mean cross validation score untuk setiap K			
	$k = 5$	$k = 10$	$k = 15$	$k = 20$
70:30	0.8787	0.8751	0.8715	0.8708
80:20	0.8877	0.8876	0.8873	0.8872
90:10	0.8915	0.8915	0.8914	0.8915

Berdasarkan tabel 4.1 dapat dilihat bahwa nilai akurasi terbaik diberikan oleh model dengan perbandingan ukuran data *train* dan data *test* 90:10 dengan nilai $k = 5$, $k = 10$ dan $k = 20$ untuk *k-fold cross validation* nya. Oleh karena itu pembagian data yang digunakan untuk menghasilkan model dengan akurasi terbaik adalah 90:10 dan juga nilai k yang digunakan untuk *k-fold cross validation* adalah $k = 5$. Peneliti memilih menggunakan $k = 5$ karena menurut (Anguita et al., 2012), nilai k yang lebih kecil lebih disarankan, alasannya adalah bahwa dengan nilai k yang lebih kecil, model akan memiliki lebih sedikit lipatan, yang dapat membantu untuk mengurangi variabilitas dan meningkatkan interpretabilitas hasil validasi silang

Perlu diketahui bahwa *k-fold cross validation* pada tahap ini memiliki fungsi sebagai alat uji untuk menghasilkan nilai akurasi terbaik sehingga dapat terlihat kombinasi ukuran *split* data dan nilai k yang paling baik.

4.3. Proses Klasifikasi *Naïve Bayes*

Pada tahap ini proses pembelajaran model *naïve bayes* akan menggunakan ukuran pembagian data 90:10. Ada beberapa tahap penting dalam proses klasifikasi *Naïve bayes*, yaitu sebagai berikut:

1. Melakukan *Split* Data

Data yang sudah dikumpulkan kemudian dibagi menjadi data *train* dan data *test*, maka setelah data dibagi dengan perbandingan 90% data *train* dan 10% data *test* diperoleh data *train* sebanyak 351 data dan data *test* sebanyak 39 data.

2. Menghitung Probabilitas Prior

Probabilitas prior adalah probabilitas suatu kejadian sebelum adanya tambahan informasi. Dalam data *train* yang terdiri dari 351 data, terdapat 129 data yang termasuk dalam kelas miskin dan 222 data dalam kelas non-miskin. Untuk menghitung probabilitas prior kelas miskin dan non-miskin, dapat dilakukan dengan cara berikut:

- Probabilitas Kelas Miskin (B_1)

$$P(B_1) = \frac{\text{Jumlah data miskin}}{\text{jumlah seluruh data}} = \frac{126}{351} = 0.3589$$

- Probabilitas Kelas Non-Miskin (B_2)

$$P(B_2) = \frac{\text{Jumlah data non - miskin}}{\text{jumlah seluruh data}} = \frac{225}{351} = 0.6410$$

3. Menghitung Probabilitas Likelihood

Probabilitas likelihood adalah probabilitas bahwa variabel-variabel tertentu muncul dalam kelas miskin atau non-miskin. Maka dihitung probabilitas untuk setiap atribut A pada data latih, yaitu sebagai berikut:

- Probabilitas $A_1 = 1$ Dalam Kelas Miskin ($P(A_1|B_1)$)

$$P(A_1|B_1) = \frac{\text{Jumlah data } A_1 = 1 \text{ pada kelas miskin}}{\text{jumlah data miskin}}$$

$$= \frac{41}{126} = 0.3254$$

- Probabilitas $A_1 = 1$ Dalam Kelas Non-Miskin ($P(A_1|B_2)$)

$$P(A_1|B_2) = \frac{\text{Jumlah data } A_1 = 1 \text{ pada kelas non - miskin}}{\text{jumlah data non - miskin}}$$

$$= \frac{78}{225} = 0.3467$$

di mana:

A_1 = Variabel jumlah tanggungan

B_1 = Kelas miskin

B_2 = Kelas non-miskin

$P(A_1|B_1)$ = Peluang kemunculan nilai 1 pada A_1 di dalam kelas miskin

$P(A_1|B_2)$ = Peluang kemunculan nilai 1 pada A_1 di dalam kelas non-miskin

Setelah semua data yang ada dihitung menggunakan cara diatas, maka hasil untuk perhitungan probabilitas likelihood variabel jumlah tanggungan dapat dilihat pada tabel 4.2.

Tabel 4. 3. Probabilitas Likelihood Variabel Jumlah Tanggungan

Jumlah Tanggungan (A_1)	Total Kejadian		Probabilitas ($P(A_1 B_i)$)	
	Miskin (B_1)	Non-Miskin (B_2)	Miskin (B_1)	Non-Miskin (B_2)
1	41	78	0.3254	0.3467
2	37	73	0.2937	0.3244
3	43	58	0.3413	0.2578
4	5	16	0.0397	0.0711
Total	126	225	1	1

Dari tabel 4.2 terlihat bahwa data $A_1 = 4$ memiliki nilai probabilitas terendah pada data dengan kelas miskin dan non-miskin, yaitu sebesar 0.0397 pada kelas miskin dan 0.0711 pada kelas non-miskin. Kemudian data $A_1 = 3$ memiliki nilai probabilitas tertinggi pada data dengan kelas miskin yaitu sebesar 0.3413. Sementara itu pada data dengan kelas non-miskin nilai probabilitas tertinggi dihasilkan oleh data $A_1 = 1$ dengan nilai probabilitas sebesar 0.3467.

Sehingga dengan menggunakan cara yang sama seperti pada variabel jumlah tanggungan didapat hasil perhitungan probabilitas likelihood untuk variabel lain, yaitu sebagai berikut:

Tabel 4. 4. Hasil Probabilitas Likelihood Variabel Pendidikan Terakhir

Pendidikan Terakhir (A_2)	Total Kejadian		Probabilitas ($P(A_2 B_i)$)	
	Miskin (B_1)	Non-Miskin (B_2)	Miskin (B_1)	Non-Miskin (B_2)
Tamat SD/Sederajat	41	10	0.3254	0.0444
SLTA/Sederajat	61	160	0.4841	0.7111
Strata-II	0	7	0	0.0311
SLTP/Sederajat	24	15	0.1905	0.0667
Diploma-IV/Strata-I	0	32	0	0.1422
Diploma-III	0	1	0	0.0044
Total	126	225	1	1

Dari tabel 4.3 terlihat bahwa pada variabel pendidikan terakhir terdapat 3 kategori yang memiliki nilai probabilitas 0 pada data dengan kelas miskin, yaitu Strata-II, Diploma-IV/Strata-I, dan Diploma III. Lalu nilai probabilitas tertinggi pada data dengan kelas miskin dihasilkan oleh kategori SLTA/Sederajat dengan nilai probabilitas 0.4841. Kemudian untuk data dengan kelas non-miskin nilai probabilitas tertinggi dihasilkan oleh kategori SLTA/Sederajat dengan nilai probabilitas sebesar 0.7111 dan nilai probabilitas terendah dihasilkan oleh kategori Diploma-III dengan nilai probabilitas sebesar 0.0044.

Tabel 4. 5. Hasil Probabilitas Likelihood Variabel Pekerjaan

Pekerjaan (A_3)	Total Kejadian		Probabilitas ($P(A_3 B_i)$)	
	Miskin (B_1)	Non-Miskin (B_2)	Miskin (B_1)	Non-Miskin (B_2)
Wiraswasta	77	60	0.6111	0.2667
Karyawan Swasta	49	123	0.3889	0.5467
Guru	0	9	0	0.04
Pensiunan	0	20	0	0.0889
Pegawai Negeri Sipil	0	7	0	0.0311

Karyawan BUMN	0	6	0	0.0267
Total	126	225	1	1

Dari tabel 4.4 terlihat bahwa hanya ada dua kategori pekerjaan yang muncul pada data dengan kelas miskin, yaitu wiraswasta dan karyawan swasta dengan masing-masing nilai probabilitas yang dihasilkan sebesar 0.6111 untuk kategori wiraswasta dan 0.3889 untuk kategori karyawan swasta. Kemudian untuk data dengan kelas non-miskin nilai probabilitas tertinggi dihasilkan oleh kategori karyawan swasta dengan nilai probabilitas sebesar 0.5467 dan nilai probabilitas terendah dihasilkan oleh kategori karyawan BUMN dengan nilai probabilitas sebesar 0.0267.

Tabel 4. 6. Hasil Probabilitas Likelihood Variabel Penghasilan

Penghasilan (A_4)	Total Kejadian		Probabilitas ($P(A_4 B_i)$)	
	Miskin (B_1)	Non-Miskin (B_2)	Miskin (B_1)	Non-Miskin (B_2)
500,000	21	0	0.1667	0
700,000	13	0	0.1031	0
800,000	18	0	0.1249	0
⋮	⋮	⋮	⋮	⋮
5,500,000	0	1	0	0.0044
5,600,000	0	4	0	0.0178
6,000,000	0	2	0	0.0089
8,000,000	0	4	0	0.0178
Total	126	225	1	1

Dari tabel 4.5 terlihat bahwa pada data dengan kelas miskin, nilai probabilitas tertinggi dihasilkan oleh kategori penghasilan 2,000,000 dengan nilai probabilitas sebesar 0.1905. Kemudian untuk data dengan kelas non-miskin nilai probabilitas tertinggi dihasilkan oleh kategori penghasilan 2,000,000 dengan nilai probabilitas sebesar 0.1333.

Tabel 4. 7. Hasil Probabilitas Likelihood Variabel Jumlah Mobil

Jumlah Mobil (A_5)	Total Kejadian		Probabilitas ($P(A_5 B_i)$)	
	Miskin (B_1)	Non-Miskin (B_2)	Miskin (B_1)	Non-Miskin (B_2)
0	126	197	1	0.8755
1	0	28	0	0.1244
Total	126	225	1	1

Dari tabel 4.7 terlihat bahwa pada variabel jumlah mobil hanya ada satu kategori jumlah mobil yang muncul yaitu kategori tidak memiliki mobil atau $A_5 = 0$ sehingga nilai probabilitasnya adalah 1. Sementara itu pada data dengan kelas non-miskin ada dua kategori jumlah mobil yang muncul yaitu $A_5 = 0$ dan $A_5 = 1$ dengan nilai probabilitas sebesar 0.8755 untuk kategori $A_5 = 0$ dan nilai probabilitas sebesar 0.1244 untuk kategori $A_5 = 1$.

Tabel 4. 8. Hasil Probabilitas Likelihood Variabel Jumlah Motor

Jumlah Sepeda Motor (A_6)	Total Kejadian		Probabilitas ($P(A_6 B_i)$)	
	Miskin (B_1)	Non-Miskin (B_2)	Miskin (B_1)	Non-Miskin (B_2)
0	47	0	0.3730	0
1	79	108	0.6270	0.48
2	0	110	0	0.4889
3	0	7	0	0.0311
Total	126	225	1	1

Dari tabel 4.8 terlihat bahwa pada variabel jumlah sepeda motor hanya ada dua kategori yang muncul pada data dengan kelas miskin, yaitu $A_6 = 0$ dan $A_6 = 1$ dengan nilai probabilitas sebesar 0.3730 untuk kategori $A_6 = 0$ dan nilai probabilitas sebesar 0.6270 untuk kategori $A_6 = 1$. Kemudian untuk data dengan kelas non-miskin memiliki tiga kategori, yaitu $A_6 = 1$, $A_6 = 2$ dan $A_6 = 3$ dengan nilai probabilitas tertinggi dihasilkan oleh kategori $A_6 = 2$

sebesar 0.4889 dan nilai probabilitas terendah dihasilkan oleh kategori $A_6 = 3$ sebesar 0.0311.

Tabel 4. 9. Hasil Probabilitas Likelihood Variabel Status Kepemilikan Rumah

Status Kepemilikan Rumah (A_7)	Total Kejadian		Probabilitas ($P(A_7 B_i)$)	
	Miskin (B_1)	Non-Miskin (B_2)	Miskin (B_1)	Non-Miskin (B_2)
Sewa	23	19	0.1825	0.0844
Milik Sendiri	87	171	0.6905	0.76
Ikut Orang Tua	16	35	0.1270	0.1556
Total	126	225	1	1

Dari tabel 4.9 terlihat bahwa pada variabel status kepemilikan rumah untuk data dengan kelas miskin nilai probabilitas tertinggi dihasilkan oleh kategori milik sendiri dengan nilai probabilitas sebesar 0.6905 dan nilai probabilitas terendah dihasilkan oleh kategori ikut orang tua dengan nilai probabilitas sebesar 0.1270. Kemudian untuk data dengan kelas non-miskin nilai probabilitas tertinggi dihasilkan oleh kategori milik sendiri dengan nilai probabilitas sebesar 0.76 dan nilai probabilitas terendah dihasilkan oleh kategori sewa dengan nilai probabilitas sebesar 0.0844.

Tabel 4. 10. Hasil Probabilitas Likelihood Variabel Kondisi Rumah

Kondisi Rumah (A_8)	Total Kejadian		Probabilitas ($P(A_8 B_i)$)	
	Miskin (B_1)	Non-Miskin (B_2)	Miskin (B_1)	Non-Miskin (B_2)
Permanen	34	214	0.2698	0.9511
Semi Permanen	92	11	0.7302	0.0489
Total	126	225	1	1

Dari tabel 4.10 terlihat bahwa pada variabel kondisi rumah untuk data dengan kelas miskin nilai probabilitas tertinggi dihasilkan oleh kategori semi permanen dengan nilai probabilitas sebesar 0.7302 dan nilai probabilitas terendah dihasilkan oleh kategori permanen dengan nilai probabilitas sebesar

0.2698. Sementara itu pada data dengan kategori non-miskin nilai probabilitas tertinggi dihasilkan oleh kategori permanen dengan nilai probabilitas sebesar 0.9511 dan nilai probabilitas terendah dihasilkan oleh kategori semi permanen dengan nilai probabilitas sebesar 0.0489.

4. Menghitung Probabilitas Posterior

Langkah terakhir dari proses klasifikasi Naïve Bayes adalah menghitung probabilitas posterior. Probabilitas posterior adalah probabilitas kelas miskin dan non-miskin berdasarkan probabilitas prior dan likelihood yang sudah di hitung sebelumnya.

$$P(B_j|A_i) = \frac{P(A_1|B_j) \times \dots \times P(A_i|B_j) \times P(B_j)}{\sum_{k=1}^i P(A_1|B_k) \times \dots \times P(A_i|B_k) \times P(B_k)}$$

Maka hasil dari perhitungan probabilitas posterior dengan menggunakan data uji, yaitu sebagai berikut:

Jumlah tanggungan = 2; Pendidikan terakhir = SLTA/Sederajat; Pekerjaan = Karyawan Swasta; Penghasilan = 2,500,000; Jumlah Mobil = 0; Jumlah motor = 1; Status kepemilikan rumah = Milik Sendiri; Kondisi rumah = Permanen. Maka untuk nilai probabilitas prior dan likelihood dapat dilihat dari langkah sebelumnya.

- Probabilitas prior
 - a. Kelas miskin (B_1) = 0.3589
 - b. Kelas non-miskin (B_2) = 0.6411
- Probabilitas likelihood
 - a. Jumlah tanggungan (A_1)
 - $P(A_1|B_1) = 0.2937$
 - $P(A_1|B_2) = 0.3244$
 - b. Pendidikan Terakhir (A_2)
 - $P(A_2|B_1) = 0.4841$
 - $P(A_2|B_2) = 0.7111$
 - c. Pekerjaan (A_3)

- $P(A_3|B_1) = 0.3889$
- $P(A_3|B_2) = 0.5467$
- d. Penghasilan (A_4)
 - $P(A_4|B_1) = 0.1190$
 - $P(A_4|B_2) = 0.08$
- e. Jumlah Mobil (A_5)
 - $P(A_5|B_1) = 1$
 - $P(A_5|B_2) = 0.8755$
- f. Jumlah Motor (A_6)
 - $P(A_6|B_1) = 0.6270$
 - $P(A_6|B_2) = 0.48$
- g. Status Kepemilikan Rumah (A_7)
 - $P(A_7|B_1) = 0.6965$
 - $P(A_7|B_2) = 0.76$
- h. Kondisi Rumah (A_8)
 - $P(A_8|B_1) = 0.2698$
 - $P(A_8|B_2) = 0.9511$

- Probabilitas Posterior

- a. Probabilitas posterior kelas miskin

$$\begin{aligned}
 P(B_1|A_i) &= \frac{P(A_1|B_1) \times \dots \times P(A_i|B_1) \times P(B_1)}{(P(A_i|B_1) \times P(B_1)) + (P(A_i|B_2) \times P(B_2))} \\
 &= \frac{0.000774326 \times 0.3589}{(0.000774326 \times 0.3589) + (0.003064985 \times 0.6411)} \\
 &= \frac{0.0002779056014}{0.0002779056014 + 0.0019649618835} \\
 &= 0.1239
 \end{aligned}$$

- b. Probabilitas posterior kelas non-miskin

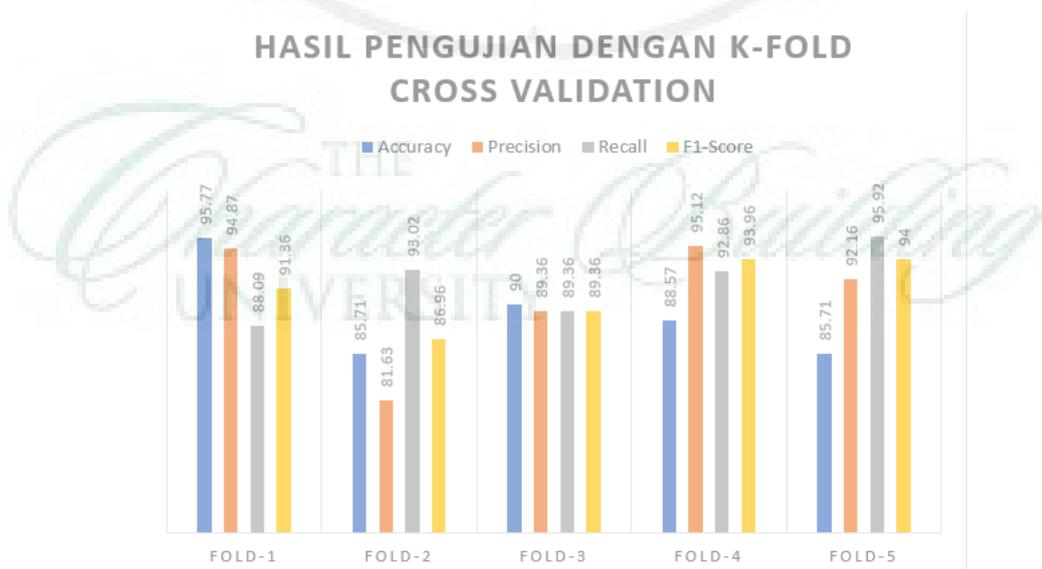
$$P(B_2|A_i) = \frac{P(A_1|B_2) \times \dots \times P(A_i|B_2) \times P(B_2)}{(P(A_i|B_1) \times P(B_1)) + (P(A_i|B_2) \times P(B_2))}$$

$$\begin{aligned}
 &= \frac{0.003064985 \times 0.6411}{(0.000774326 \times 0.3589) + (0.003064985 \times 0.6411)} \\
 &= \frac{0.0019649618835}{0.0002779056014 + 0.0019649618835} \\
 &= 0.8761
 \end{aligned}$$

Berdasarkan hasil perhitungan probabilitas posterior di atas dapat disimpulkan bahwa nilai $P(B_1|A_i) < P(B_2|A_i)$. Maka data uji tersebut diklasifikasikan ke dalam kelas **Non-Miskin**.

4.4. Evaluasi Model

Setelah mengetahui tahapan untuk menerapkan algoritma *Naïve Bayes Classifier*, maka langkah selanjutnya adalah penerapan model klasifikasi dengan menggunakan bahasa pemrograman Python dan melalui bantuan *software Visual Studio Code*. Kemudian model dievaluasi dengan menggunakan *k-fold cross validation* untuk melihat akurasi yang dihasilkan. *K-Fold Cross Validation* berfungsi untuk memperoleh akurasi yang lebih stabil dan optimal sehingga mendapatkan gambaran yang lebih baik mengenai kinerja dari model yang sudah dibangun (Tempola et al., 2018). Untuk hasil evaluasi menggunakan *K-Fold Cross Validation* dapat dilihat pada tabel 4.11 berikut.



Gambar 4. 10. Diagram Hasil Pengujian dengan *K-Fold Cross Validation*

Dari gambar 4.10 terlihat bahwa ada akurasi yang dihasilkan pada setiap fold berbeda, maka langkah terakhir dari evaluasi dengan menggunakan *k-fold cross validation* yaitu dengan menghitung nilai rata-rata sehingga didapatkan *mean score cross validation*, maka didapatkan nilai akurasi sebesar 89.15%, *recall* sebesar 91.58%, presisi sebesar 90.63% dan *f1-score* sebesar 91.13%. Menurut (Putri et al., 2021), apabila akurasi model sudah melebihi nilai 80% maka sudah terbilang cukup tinggi sehingga dapat dikategorikan “Baik”. Maka dengan demikian model klasifikasi yang dibangun sudah dapat dipercaya.

4.5. Penerapan Sistem

Sistem dirancang berbasis web dengan menggunakan bahasa pemrograman Python, Javascript, dan HTML. Sistem ini merupakan sistem klasifikasi penduduk miskin yang berfungsi untuk melakukan pengelompokkan data yang lebih akurat dan lebih efisien karena dibangun dengan menerapkan algoritma *Naïve Bayes Classifier* dan sudah dilatih dengan menggunakan data penduduk yang diperoleh dari Kantor Kepala Desa Tembung. Kemudian sistem akan divalidasi untuk memastikan bahwa web sudah berjalan dengan baik dan tidak ada kesalahan. Sistem divalidasi dengan cara mencoba semua fitur yang ada, lalu untuk memastikan akurasi yang diberikan sudah baik, sistem diuji dengan menginput data sebanyak 10% dari jumlah keseluruhan data sampel atau sebanyak 39 data uji, kemudian akurasi dihitung menggunakan persamaan 11 berikut (Pratiwi et al., 2021):

$$Akurasi = \frac{Klasifikasi\ benar}{Jumlah\ data\ uji} \times 100\% \quad (11)$$

Maka setelah melakukan pengujian dengan menginput 39 data uji kedalam sistem, rangkuman mengenai hasil klasifikasi terhadap data uji dapat dilihat pada tabel 4.10 berikut:

Tabel 4. 11. Rangkuman Hasil Klasifikasi Pada Sistem

A1	A2	...	A7	A8	Label Asli	Label Prediksi
1	Tamat SD/Sederajat	...	Milik Sendiri	Semi Permanen	Miskin	Miskin
2	SLTA/Sederajat	...	Milik Sendiri	Semi Permanen	Miskin	Miskin
2	SLTA/Sederajat	...	Ikut Orang Tua	Semi Permanen	Miskin	Miskin
3	SLTA/Sederajat	...	Milik Sendiri	Semi Permanen	Miskin	Miskin
4	SLTA/Sederajat	...	Milik Sendiri	Permanen	Non-Miskin	Non-Miskin
3	SLTP/Sederajat	...	Sewa	Semi Permanen	Miskin	Miskin
1	SLTA/Sederajat	...	Ikut Orang Tua	Permanen	Non-Miskin	Non-Miskin
3	Strata-II	...	Milik Sendiri	Permanen	Non-Miskin	Non-Miskin
1	Tamat SD/Sederajat	...	Milik Sendiri	Permanen	Non-Miskin	Miskin
1	Tamat SD/Sederajat	...	Milik Sendiri	Semi Permanen	Miskin	Miskin
...
1	SLTP/Sederajat	...	Milik Sendiri	Permanen	Miskin	Miskin
2	SLTA/Sederajat	...	Sewa	Permanen	Non-Miskin	Non-Miskin
2	SLTA/Sederajat	...	Ikut Orang Tua	Permanen	Non-Miskin	Non-Miskin
3	SLTA/Sederajat	...	Milik Sendiri	Semi Permanen	Non-Miskin	Miskin
2	SLTA/Sederajat	...	Milik Sendiri	Permanen	Non-Miskin	Non-Miskin
2	Diploma IV/Strata I	...	Milik Sendiri	Permanen	Non-Miskin	Non-Miskin

Setelah semua data diinput, tercatat data uji yang diklasifikasikan dengan benar ada sebanyak 35 data dan data uji yang diklasifikasikan dengan salah ada sebanyak 4 data. Maka dengan menggunakan persamaan 11 didapatkan nilai akurasi sebesar 89.74%. Adapun sistem yang sudah selesai dirancang adalah sebagai berikut:

1. Halaman Login

Halaman ini merupakan halaman yang muncul pertama kali saat pengguna mengakses *website*.



Gambar 4. 11. Halaman Login

2. Halaman Dashboard

Halaman dashboard merupakan halaman yang muncul pertama kali saat pengguna telah berhasil login. Halaman ini menampilkan data kepala keluarga yang sudah di klasifikasi.

Nama Kepala Keluarga	No. Kartu Keluarga	Alamat	Jumlah Tanggungan	Pendidikan Terakhir	Pekerjaan	Penghasilan	Jumlah Mobil	Jumlah Motor	St
P0001	KK0001	A0001	1	TAMAT SD/SEDERAJAT	WIRASWASTA	500000	0	0	M
P0002	KK0002	A0002	2	TAMAT SD/SEDERAJAT	KARYAWAN SWASTA	2000000	0	1	M
P0003	KK0003	A0003	4	TAMAT SD/SEDERAJAT	WIRASWASTA	2000000	0	1	M
P0004	KK0004	A0004	3	SLTA/SEDERAJAT	WIRASWASTA	1800000	0	1	M
P0005	KK0005	A0005	2	SLTA/SEDERAJAT	KARYAWAN SWASTA	3200000	0	2	SE
P0006	KK0006	A0006	2	SLTA/SEDERAJAT	KARYAWAN SWASTA	1800000	0	1	M
P0007	KK0007	A0007	1	SLTA/SEDERAJAT	KARYAWAN SWASTA	3100000	0	1	M
P0008	KK0008	A0008	1	SLTA/SEDERAJAT	WIRASWASTA	1800000	0	2	M
P0009	KK0009	A0009	1	SLTP/SEDERAJAT	WIRASWASTA	500000	0	1	M
P0010	KK0010	A0010	4	TAMAT SD/SEDERAJAT	WIRASWASTA	2000000	0	1	M
P0011	KK0011	A0011	3	SLTA/SEDERAJAT	KARYAWAN SWASTA	2800000	0	1	M

Gambar 4. 12. Halaman Dashboard

3. Halaman Klasifikasi

Halaman klasifikasi merupakan halaman yang berfungsi untuk menginput data baru yang kemudian diklasifikasikan dan diinput kedalam data kepala keluarga.

Sistem Klasifikasi Kemiskinan

Dashboard
Klasifikasi
Admin
Keluar

Klasifikasi Data Rumah Tangga

Nama Lengkap Kepala Keluarga
Gabriel Christian

Nomor Kartu Keluarga
123786321871123

Alamat
Jl. Pancasila, Kec. Percut Sei Tuan, Kabupaten Deli Serdang, Sumatera Utara

Jumlah Tanggungan
2

Pendidikan Terakhir
Diploma-IV/Strata-I

Pekerjaan
Karyawan Swasta

Penghasilan Perbulan
3000000

Jumlah Mobil
0

Jumlah Motor

Gambar 4. 13. Halaman Klasifikasi

4. Halaman Admin

Halaman admin merupakan halaman yang digunakan untuk melakukan manajemen akun admin. Pada halaman ini admin dapat menambahkan, mengedit, dan menghapus akun.

Sistem Klasifikasi Kemiskinan

Dashboard
Klasifikasi
Admin
Keluar

Daftar Admin Yang Terdaftar

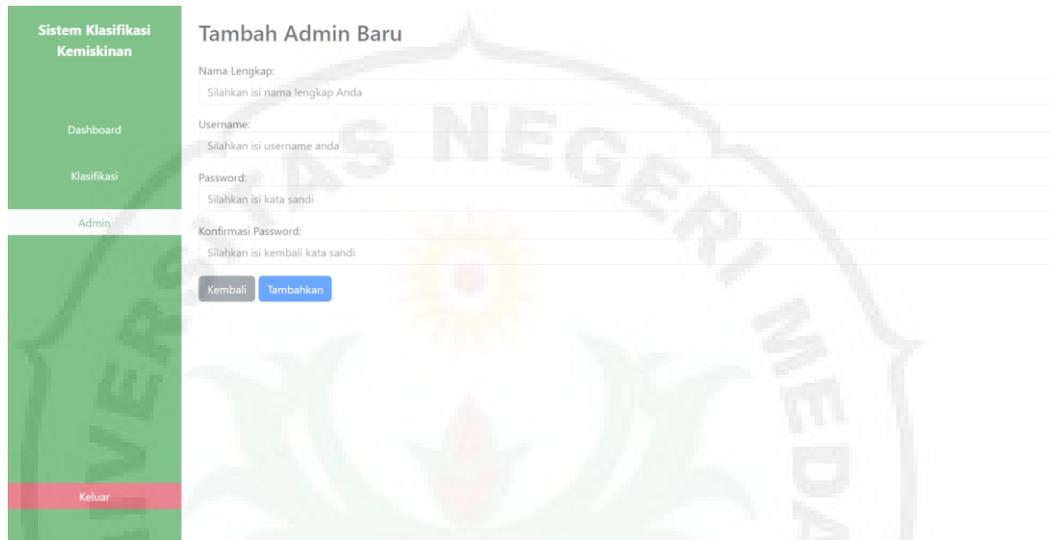
Tambah Admin

Nama Lengkap	Username	Opsi
Gabriel Christian	admin	Edit Hapus

Gambar 4. 14. Halaman Admin

5. Halaman Tambah Admin

Halaman tambah admin merupakan halaman yang berfungsi untuk menambahkan akun admin baru supaya dapat login ke dalam aplikasi.



Sistem Klasifikasi Kemiskinan

Tambah Admin Baru

Nama Lengkap:
Silahkan isi nama lengkap Anda

Username:
Silahkan isi username anda

Password:
Silahkan isi kata sandi

Konfirmasi Password:
Silahkan isi kembali kata sandi

Kembali Tambahkan

Dashboard

Klasifikasi

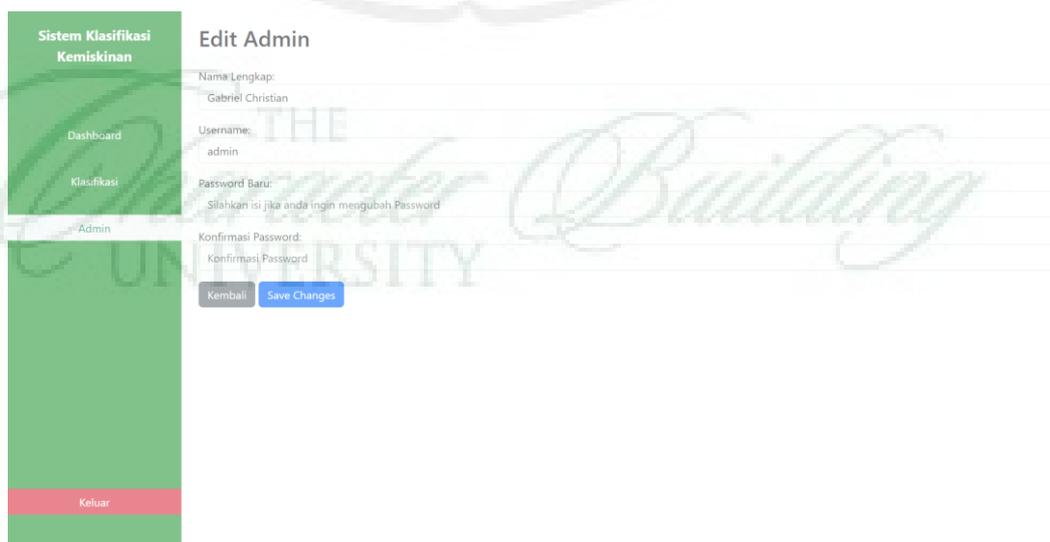
Admin

Keluar

Gambar 4. 15. Halaman Tambah Admin

6. Halaman Edit Admin

Halaman edit admin merupakan halaman yang berfungsi untuk melakukan edit data yang berkaitan dengan akun terdaftar.



Sistem Klasifikasi Kemiskinan

Edit Admin

Nama Lengkap:
Gabriel Christian

Username:
admin

Password Baru:
Silahkan isi jika anda ingin mengubah Password

Konfirmasi Password:
Konfirmasi Password

Kembali Save Changes

Dashboard

Klasifikasi

Admin

Keluar

Gambar 4. 16. Halaman Edit Admin