

## Penentuan Kelayakan Masyarakat Miskin Penerima Bantuan Menggunakan Metode Naïve Bayes (Studi Kasus: Kabupaten Penajam Paser Utara)

Nur Madia <sup>(1)</sup>, Anindita Septiarini <sup>(2)\*</sup>, Heliza Rahmania Hatta <sup>(3)</sup>, Hamdani Hamdani <sup>(3)</sup>,  
Masna Wati <sup>(3)</sup>

Informatika, Fakultas Teknik, Universitas Mulawarman, Samarinda  
e-mail : nmndia18@gmail.com, {anindita,heliza,hamdani}@unmul.ac.id,  
masnawati@fkti.unmul.ac.id.

\* Penulis korespondensi.

Artikel ini diajukan 7 November 2022, direvisi 27 Desember 2022, diterima 2 Januari 2023, dan dipublikasikan 30 Januari 2023.

### Abstract

*Contents Poverty is the inability to meet the necessities of life, such as food, clothing, and shelter. The poor have an average monthly per capita expenditure below the poverty line. The case of poverty in Indonesia is still unresolved; the Government continues to try to give the best to the entire community so that the problem of poverty can at least continue to decrease. One form of government concern for the poor is the assistance program provided to the poor. This study will classify based on data from the North Penajam Paser (PPU) community obtained from the results of the National Socio-Economic Survey (Susenas) to know how the Naïve Bayes method is in determining the eligibility of the poor recipients of assistance. Based on the research that has been carried out, a system for determining the poor recipients of assistance is produced, where the test results get the highest accuracy in the third scenario, namely 60% or 328 training data and 40% or 218 test data, where the accuracy obtained is 77.98%.*

**Keywords:** Classification, Poor Society, Laplace Correction, Data Mining, Naïve Bayes

### Abstrak

Kemiskinan adalah ketidakmampuan dalam memenuhi kebutuhan hidup seperti makanan, pakaian, dan tempat tinggal. Penduduk miskin adalah penduduk yang memiliki rata-rata pengeluaran perkapita per-bulan di bawah garis kemiskinan. Kasus kemiskinan di Indonesia hingga saat ini masih belum terselesaikan, pemerintah terus berusaha memberikan yang terbaik kepada seluruh masyarakat agar permasalahan kemiskinan setidaknya bisa terus berkurang. Salah satu bentuk kepedulian pemerintah pada masyarakat miskin adalah program bantuan yang diberikan kepada masyarakat miskin. Penelitian ini melakukan klasifikasi berdasarkan data masyarakat Penajam Paser Utara (PPU) yang didapatkan dari hasil Survey Sosial Ekonomi Nasional (Susenas). Klasifikasi dilakukan untuk menentukan kelayakan masyarakat miskin dalam penerima bantuan yang diterapkan menggunakan metode Naïve Bayes. Sejumlah 547 data digunakan untuk evaluasi kinerja dari metode Naïve Bayes. Data tersebut dibedakan menjadi dua sebagai data latih dan data uji yaitu 60% (328 data latih) dan 40% (219 data uji). Berdasarkan evaluasi yang dilakukan, nilai akurasi yang dihasilkan mencapai 77,98%.

**Kata Kunci:** Klasifikasi, Masyarakat Miskin, Laplace Correction, Data Mining, Naïve Bayes

## 1. PENDAHULUAN

Kemiskinan merupakan salah satu permasalahan sosial yang belum terselesaikan di Indonesia. Berdasarkan data dari Badan Pusat Statistik (BPS) jumlah masyarakat miskin di Indonesia tahun 2021 mengalami kenaikan dari tahun sebelumnya. Salah satu daerah di Indonesia yang juga mengalami kenaikan tersebut adalah Kabupaten Penajam Paser Utara. Persentase tingkat kemiskinan pada tahun 2021 di Penajam mengalami kenaikan sebanyak 0,25% dari tahun sebelumnya (Badan Pusat Statistik, 2021). Kemiskinan merupakan masalah yang dipengaruhi oleh faktor-faktor yang saling berkaitan, seperti tingkat pendapatan masyarakat, pengangguran,



kesehatan, pendidikan, akses terhadap barang dan jasa, lokasi, geografis, gender, dan lokasi lingkungan (Novriansyah, 2018).

Pemerintah mendirikan program pengentasan dalam mengatasi permasalahan kemiskinan di Indonesia. Program utama yang dilakukan adalah kelompok program yang bertujuan untuk mengurangi beban hidup dalam memenuhi kebutuhan pangan, kesehatan, dan pendidikan. Contoh program seperti pendistribusian beras miskin (Raskin dan BPNT), pemberian Jaminan Kesehatan Masyarakat (Jamkesmas), pemberian bantuan keuangan (BLT dan PKH), bantuan pendidikan bagi siswa miskin (BSM dan PIP), dan Bantuan/Subsidi Pemerintah Daerah. Program tersebut secara langsung berkaitan dengan terwujudnya hak asasi manusia (Umami, 2013). Bantuan yang diberikan melewati suatu proses sehingga bantuan tersalurkan dengan tepat sasaran kepada masyarakat miskin. Proses yang dapat dilakukan dalam menentukan masyarakat miskin penerima bantuan yaitu dengan melakukan klasifikasi masyarakat miskin menggunakan metode *machine learning* seperti pada penelitian Aji (2019), Arifando et al. (2017), Fitriani (2020), Kurnia et al. (2019), Purnama et al. (2020), dan Rihanah & Fatmawati (2021).

Klasifikasi menggunakan metode Naïve Bayes telah banyak digunakan dalam penelitian terkait penentuan masyarakat miskin dan penerima bantuan. Naïve Bayes merupakan teknik prediksi probabilistik sederhana yang berdasar pada penerapan teorema Bayes dengan asumsi independensi yang kuat (Aji, 2019). Metode ini memiliki sifat yang efektif dan cepat dalam mengolah data berjumlah besar. Kemampuan tersebut membuat metode ini sering digunakan pada aplikasi seperti *spam filtering* dan deteksi anomali di jaringan komputer (Kurniawan, 2020). Metode Naïve Bayes juga memiliki kemampuan yang baik dari metode *data mining* lainnya seperti *Support Vector Machine* dalam melakukan klasifikasi (Maarif, 2016).

Penelitian sebelumnya terkait klasifikasi masyarakat miskin dan penerima bantuan telah dilakukan oleh Putri et al. (2021). Pada penelitian tersebut dilakukan klasifikasi untuk rumah tangga miskin di Provinsi Papua menggunakan metode Naïve Bayes. Variabel yang digunakan pada penelitian tersebut yaitu, jenis kelamin, pendidikan KRT, lapangan usaha KRT, jenis atap terluas, jenis dinding terluas, jenis lantai terluas, sumber air minum, sumber penerangan, dan bahan bakar untuk memasak. Hasil diperoleh nilai akurasi sebesar 80%. Kemudian pada penelitian lainnya yang dilakukan oleh Annur (2018) melakukan klasifikasi penduduk miskin di Kecamatan Tibawa. Variabel yang digunakan yaitu umur, pendidikan, pekerjaan, penghasilan, tanggungan, dan status (kawin/belum kawin). Hasilnya didapatkan nilai akurasi sebesar 73%. Selanjutnya penelitian yang telah dilakukan oleh Nurmawati et al. (2021) melakukan klasifikasi masyarakat miskin menggunakan metode Naïve Bayes dan variabel yang digunakan yaitu jenis kelamin, peserta pkh, jumlah anggota rumah tangga, status kepemilikan rumah, fasilitas rumah, luas lantai rumah, dan fasilitas elektronik. Berdasarkan hasil penelitian yang telah dilakukan diperoleh nilai akurasi sebesar 96,63%. Kesimpulan yang diperoleh yaitu metode Naïve Bayes memiliki kemampuan yang baik untuk melakukan klasifikasi masyarakat miskin.

Pada penelitian ini metode Naïve Bayes diterapkan untuk melakukan klasifikasi kelayakan masyarakat miskin penerima bantuan di wilayah Kabupaten Penajam Paser Utara. Perbedaan dengan penelitian sebelumnya terletak pada jenis variabel yang digunakan. Pada penelitian ini jumlah variabel yang digunakan lebih banyak dari penelitian sebelumnya, di mana variabel tersebut dapat menggambarkan karakteristik dari masyarakat miskin. Data yang digunakan pada penelitian ini merupakan data yang diperoleh dari Survey Sosial Ekonomi Nasional (Susenas) KOR RT bulan Maret Tahun 2021 pada wilayah Kabupaten Penajam Paser Utara. Data yang dikumpulkan terdiri dari 546 data dengan jumlah data masyarakat yang layak menerima bantuan sebanyak 172 data dan masyarakat yang tidak menerima bantuan sebanyak 374 data.

## 2. METODE PENELITIAN

### 2.1 Pengumpulan Data

Data yang dikumpulkan pada penelitian ini berupa data sekunder. Data yang dikumpulkan berupa data Survei Sosial Ekonomi Nasional KOR RT Wilayah Penajam Paser Utara Bulan Maret Tahun



2021. Data didapatkan dari situs resmi BPS Indonesia. Data yang dikumpulkan terdiri dari 41 parameter yang dapat dilihat pada Tabel 1. Adapun contoh dari data yang dikumpulkan pada penelitian ini dapat dilihat pada Tabel 2. Pada Tabel 2, nilai 1, 5, 8, dan 9 merupakan jawaban untuk “ya”, “tidak”, “tidak tahu”, dan “menolak menjawab”.

**Tabel 1 Parameter Penelitian**

Variabel	Keterangan
R1 = R1701	Selama setahun terakhir, apakah khawatir tidak akan memiliki cukup makanan?
R2 = R1702	Selama setahun terakhir, apakah ada saat di mana tidak dapat menyantap makanan sehat dan bergizi?
R3 = R1703	Selama setahun terakhir, apakah hanya menyantap sedikit jenis makanan?
R5 = R1705	Selama setahun terakhir, apakah makan lebih sedikit daripada seharusnya?
R6 = R1706	Selama setahun terakhir, apakah kehabisan makanan?
R7 = R1707	Selama setahun terakhir, apakah merasa lapar tapi tidak makan?
R8 = R1708	Selama setahun terakhir, apakah tidak makan seharian?
R9 = R1801	Berapa jumlah keluarga yang tinggal di dalam bangunan sensus/rumah ini?
R10 = R1802	Apakah status kepemilikan bangunan tempat tinggal yang ditempati?
R11 = R1803	Apa jenis bukti kepemilikan tanah bangunan tempat tinggal ini?
R12 = R1804	Berapa luas lantai rumah bangunan tempat tinggal?
R13 = R1805	Apakah KRT/pasangannya/anaknya memiliki rumah lain, selain rumah yang ditempati saat ini?
R14 = R1806	Apakah bahan bangunan utama atap rumah terluas?
R15 = R1807	Apakah bahan bangunan utama dinding rumah terluas?
R16 = R1808	Apakah bahan bangunan utama lantai rumah terluas?
R17 = R1809A	Apakah memiliki fasilitas tempat buang air besar?
R18 = R1809B	Apakah jenis kloset yang digunakan?
R19 = R1809C	Di manakah tempat pembuangan akhir tinja?
R20 = R1810A	Apa sumber air utama yang digunakan untuk minum?
R21 = R1812	Apakah pernah mengalami kekurangan air minum selama minimal 24 jam?
R22 = R1814A	Apakah sumber air utama untuk mandi/cuci/dll?
R23 = R1814B	Berapa jarak ke tempat penampungan limbah/kotoran/tinja terdekat?
R24 = R1816	Apakah sumber utama penerangan rumah tangga ini?
R25 = R1817	Apakah jenis bahan bakar utama yang digunakan untuk memasak?
R26 = R2001A	Apakah memiliki tabung gas 5,5 kg atau lebih?
R27 = R2001B	Apakah memiliki lemari es/kulkas?
R28 = R2001C	Apakah memiliki AC?
R29 = R2001D	Apakah memiliki pemanas air (water heater)?
R30 = R2001E	Apakah memiliki telepon rumah (PSTN)?
R31 = R2001F	Apakah memiliki komputer/laptop?
R32 = R2001G	Apakah memiliki emas/perhiasan (minimal 10 gram)?
R33 = R2001H	Apakah memiliki sepeda motor?
R34 = R2001I	Apakah memiliki perahu?
R35 = R2001J	Apakah memiliki perahu motor?
R36 = R2001K	Apakah memiliki mobil?
R37 = R2001L	Apakah memiliki televisi layar datar (minimal 30 inch)?
R38 = R2001M	Apakah memiliki tanah/lahan?
R39 = R2101A	Apakah sumber terbesar pembiayaan di rumah tangga ini?
R40 = R2101C	Dari manakah sumber utama kiriman uang/barang?
R41 = R301	Banyaknya ART



Tabel 2 Data Penelitian

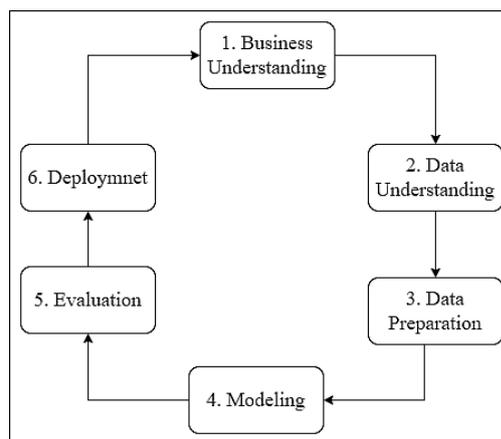
No	R1	R2	R3	R4	R5	R6	...	R38	R39	R40	R41	Label
1	5	5	5	5	5	5	...	1	1	2	3	1
2	5	5	5	5	5	5	...	1	1	2	4	0
3	5	5	5	5	5	5	...	1	1	2	3	0
4	5	5	5	5	5	5	...	5	1	2	4	0
5	5	5	5	5	5	5	...	1	1	2	5	0
6	5	5	5	5	5	5	...	5	2	2	2	0
7	5	5	5	5	5	5	...	1	1	2	1	0
8	5	5	5	5	5	5	...	1	1	2	4	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
539	5	5	5	5	5	5	...	5	1	2	4	1
540	5	5	1	1	1	1	...	5	1	2	1	0
541	1	5	1	5	5	5	...	5	1	2	6	1
542	5	5	5	5	5	5	...	1	1	2	8	1
543	5	5	5	5	5	5	...	1	1	2	3	1
544	5	5	5	5	5	5	...	1	2	2	1	1
545	5	5	5	5	5	5	...	1	2	2	1	0
546	5	5	5	5	5	5	...	1	1	2	5	0

## 2.2 Proses Pengolahan Data

*Data mining* adalah suatu proses menemukan hubungan yang berarti, pola, dan kecenderungan dengan memeriksa sekumpulan besar data yang tersimpan dalam penyimpanan dengan menggunakan teknik pengenalan pola seperti teknik statistik dan matematika (Nastuti & Harahap, 2019). *Cross Industry Standard Process for Data Mining* (CRISP-DM) merupakan salah satu metodologi dalam *data mining*. Terdapat 6 (enam) tahapan dalam metodologi CRISP-DM yang dapat dilihat pada Gambar 1.

Tahap dari metodologi CRISP-DM pada Gambar 1 dapat dijelaskan sebagai berikut (Suntoro, 2019).

- 1) *Business understanding* merupakan tahap pemahaman pada bisnis (penelitian) yang mana pada tahap tersebut menentukan tujuan bisnis (penelitian) serta kebutuhan atas tujuan yang akan dicapai.



Gambar 1 Tahapan CRISP-DM

- 2) *Data understanding* merupakan tahapan pemahaman pada data yang mana pada tahap ini peneliti mempersiapkan, mengevaluasi persiapan data, dan mengumpulkan data, selanjutnya data yang terkumpul akan dideskripsikan.



- 3) *Data preparation* disebut juga dengan data *pre-processing*, merupakan tahap membangun data ke dalam format yang diinginkan dengan mengidentifikasi, memilih, dan membersihkan data yang diperoleh.
- 4) *Modeling* adalah tahap pembuatan aplikasi dari algoritma untuk mencari, mengidentifikasi, dan menampilkan pola dari data yang akan diestimasi, prediksi, klasifikasi, clustering, atau melihat hubungan asosiatif.
- 5) *Evaluation* digunakan untuk membantu pengukuran evaluasi pada model. Pada penerapan klasifikasi, pengukuran evaluasi yang banyak digunakan adalah *accuracy*, *recall*, *precision*, *G-Mean*, *F-Measure*, dan lain sebagainya.
- 6) *Deployment*, tahapan *deployment* digunakan untuk melakukan otomatisasi model atau pengembangan aplikasi, terintegrasi dengan sistem informasi manajemen atau operasional yang ada.

### 2.3 Penggunaan Algoritma Naïve Bayes

Menurut Muljono et al. (2018) Naïve Bayes merupakan algoritma klasifikasi sederhana yang sering diimplementasikan untuk klasifikasi dokumen. Terdapat 2 tahapan dalam proses klasifikasi yaitu proses pelatihan dan pengujian. Proses pelatihan adalah pelatihan pada proses pembuatan model dengan menggunakan data latih yang sudah ditentukan label dari dokumen. Sedangkan pengujian merupakan proses untuk mengetahui keakuratan model dengan menggunakan data yang disebut dengan data uji. Adapun persamaan dari teorema Naïve Bayes ditunjukkan pada Pers. (1) (Natuzzuhriyyah et al., 2021).

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \quad (1)$$

Di mana  $X$  merupakan data dengan kelas yang belum diketahui dan  $H$  adalah hipotesis data yang merupakan suatu kelas spesifik. Sementara itu,  $P(H|X)$  dan  $P(X|H)$  adalah probabilitas hipotesis  $H$  berdasarkan kondisi  $X$  (posteriori Prob.) dan probabilitas  $X$  berdasarkan kondisi tersebut. Variabel  $P(H)$  dan  $P(X)$  merupakan hipotesis  $H$  (prior Prob.) dan probabilitas dari  $X$ .

Persamaan  $P(H|X)$  dapat disederhanakan menjadi Pers. (2) dengan asumsi independensi. Penyederhanaan ini dilakukan agar perhitungan tidak sulit dilakukan karena jika semakin banyak faktor-faktor kompleks yang mempengaruhi nilai probabilitas maka semakin mustahil untuk menghitung nilai tersebut satu persatu (Suntoro, 2019).

$$P(C|F_1, \dots, F_n) = P(C) \prod_{i=1}^n P(F_i|C) \quad (2)$$

*Dataset* dengan tipe data berupa numerik dihitung dengan perhitungan gaussian (Suntoro, 2019) di mana perhitungan tersebut dapat dilihat dari Pers. (3).

$$(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp \frac{-(x-\mu)^2}{2\sigma^2} \quad (3)$$

Perhitungan rata-rata dapat dilihat pada Pers. (4).

$$\mu = \frac{\sum_i^n x_i}{n} \quad (4)$$

Perhitungan standar deviasi dapat dilihat pada Pers. (5).

$$\sigma = \sqrt{\frac{\sum_i^n (x_i - \mu)^2}{n-1}} \quad (5)$$

Langkah dari algoritma Naïve Bayes diawali dengan menyiapkan *dataset*, kemudian menghitung jumlah kelas pada *data training*. Selanjutnya menghitung jumlah kasus yang sama dengan kelas yang sama, kemudian dilakukan terhadap semua hasil sesuai dengan *data testing* yang akan



dicari kelasnya. Terakhir adalah membandingkan hasil per kelas, nilai tertinggi ditetapkan sebagai kelas baru.

## 2.4 Laplace Correction

Hasil perhitungan menggunakan metode Naïve Bayes akan memungkinkan menghasilkan nilai perhitungan 0 karena saat dilakukan pengujian tidak ditemukan atribut pada data latih. *Laplace Correction* merupakan metode yang digunakan untuk mengatasi permasalahan tersebut. Persamaan metode *Laplace Correction* dapat dilihat pada Pers. (6) (Setiawan et al., 2021).

$$P_i = \frac{p(mi) + 1}{n + k} \quad (6)$$

Di mana  $P_i$  adalah probabilitas dari  $mi$  yang merupakan jumlah sampel dalam kelas dari anggota dengan  $n$  dan  $k$  jumlah sampel dan kelas dari atribut  $mi$ .

## 2.5 Confusion Matrix

Pengujian yang dilakukan untuk mengevaluasi kinerja dari penggunaan model klasifikasi yaitu menggunakan *confusion matrix*. Menurut Gunawan et al. (2018) terdapat empat istilah yang digunakan dalam metode tersebut yaitu:

- 1) True Positive : Jumlah data hasil output benar dan kelas output positif.
- 2) True Negative : Jumlah data hasil output benar dan kelas output negatif.
- 3) False Positive : Jumlah data hasil output salah dan kelas output positif.
- 4) False Negative : Jumlah data hasil output salah dan kelas output negatif.

Metode tersebut pada umumnya digunakan dalam menghitung tingkat *accuracy* suatu model, selain itu terdapat parameter lain yang dapat dihitung menggunakan metode tersebut yaitu, *recall* dan *precision*. Tabel dari *confusion matrix* dapat dilihat pada Tabel 3.

Tabel 3 Confusion Matrix

Kelas Target	Kelas Output	
	+	-
+	True Positives (TP)	False Negatives (FN)
-	False Positives (FP)	True Negatives (TN)

*Accuracy* yaitu pengujian yang dilakukan untuk mendapatkan nilai dari kemampuan model klasifikasi secara keseluruhan. *Recall* adalah persentase keberhasilan sistem saat menentukan ulang informasi. *Precision* adalah persentase keakuratan antara informasi yang diminta oleh pengguna dan respon yang ditunjukkan oleh system. *F1-Score* merupakan alat ukur yang digunakan untuk mendapatkan nilai rata-rata harmonik dari *recall* dan *precision*. Nilai *accuracy*, *recall*, dan *f1-score* dapat dihitung dengan menggunakan rumus pada Pers. (7) sampai (10).

$$Accuracy : \frac{(TP + TN)}{(TP + FN + FP + TN)} \times 100\% \quad (7)$$

$$Recall : \frac{TP}{(TP + FN)} \times 100\% \quad (8)$$

$$Precision : \frac{TP}{(TP + FP)} \times 100\% \quad (9)$$

$$F1 - Score : \frac{(2 \times Recall \times Precision)}{(Recall + Precision)} \times 100\% \quad (10)$$



### 3. HASIL DAN PEMBAHASAN

#### 3.1 *Data Understanding*

Pada tahap *data understanding* dijelaskan proses apa saja yang dilakukan oleh peneliti dalam memahami dan mengevaluasi persiapan data. Dalam proses memahami dan mengevaluasi persiapan data pada penelitian ini peneliti melakukan seleksi variabel, pelabelan kelas, dan eksplorasi data.

##### 3.1.1 Seleksi Variabel

Data yang digunakan merupakan data Susenas Kor RT wilayah Kabupaten Penajam Paser Utara, data tersebut berisi data diri masyarakat wilayah penajam. Variabel penelitian ini menggunakan variabel yang ada pada data Susenas kemudian dilakukan seleksi berdasarkan kriteria dari masyarakat miskin. Penelitian ini menggunakan variabel dari hasil studi literatur terkait variabel dan kriteria masyarakat miskin. Karakteristik masyarakat miskin calon penerima bantuan yang digunakan pada data penelitian ini memiliki 5 kriteria yang diduga dapat menentukan kelayakan masyarakat penerima bantuan yaitu makanan, perumahan, kepemilikan barang, sumber pembiayaan, dan banyaknya ART. Adapun penjelasan mengenai kriteria tersebut adalah sebagai berikut:

- 1) Makanan (Wati & Hadi, 2016), berisi data mengenai akses KRT/Pasangan/ART terhadap makanan. Kriteria tersebut terdiri dari beberapa variabel dan atribut yang memiliki tipe data kategorik.
- 2) Perumahan (Arifando et al., 2017; Purnama et al., 2020), berisi data jumlah keluarga yang tinggal dalam rumah, kepemilikan rumah, dan kondisi rumah serta fasilitas rumah. Kriteria tersebut terdiri dari beberapa variabel dengan atribut bertipe data kategorik atau numerik.
- 3) Kepemilikan barang (Nurmayanti et al., 2021; Sugianto & Maulana, 2019), pada kriteria tersebut berisi data barang-barang yang dimiliki suatu rumah tangga seperti TV, Kulkas, AC, Pemanas Air, Telepon Rumah, Komputer/Laptop, Kendaraan, dan Aset (Tanah & Perhiasan). Kriteria tersebut terdiri dari tiga belas variabel dengan atribut yang bertipe data kategorik.
- 4) Sumber Pembiayaan (Badan Pusat Statistik, 2021), berisi data tentang sumber pembiayaan suatu rumah tangga. Kriteria tersebut terdiri dari 2 variabel dengan atribut yang bertipe data kategorial.
- 5) Banyaknya ART (Arifando et al., 2017; Kurnia et al., 2019; Sugianto & Maulana, 2019), berisi data jumlah anggota rumah tangga dalam suatu rumah tangga. Kriteria tersebut terdiri dari satu variabel dengan atribut yang bertipe data numerik.

##### 3.1.2 Pelabelan Kelas

Proses pelabelan kelas pada data yang digunakan penelitian ini didapatkan dari data bantuan yang ada pada data Susenas. Adapun jenis bantuan yang digunakan dalam penelitian ini mencakup bantuan PKH, KKS, PIP SD, PIP SMP, PIP SM, PIP kuliah, dan BPNT. Data masyarakat yang menerima bantuan pada penjelasan sebelumnya diberikan label sebagai data masyarakat yang "Layak" menerima bantuan dan data masyarakat yang tidak menerima bantuan diatas diberi label sebagai masyarakat yang "Tidak Layak" menerima bantuan.

##### 3.1.3 Eksplorasi Data

Pemahaman data dilakukan untuk mengetahui kecenderungan pusat data yang mana hal tersebut dapat berguna untuk membersihkan data yang kotor. Proses pemahaman data yang dilakukan yaitu mencari nilai kosong = 0, data bernilai tidak tahu = 8, data bernilai menolak jawab = 9, atribut yang tidak terpakai, dan modus dari setiap atribut yang ada pada masing-masing variable data yang digunakan. Hasil proses pemahaman data dalam mencari jumlah data yang kosong = 0, data bernilai tidak tahu = 8, dan data bernilai menolak jawab = 9. Atribut tidak terpakai dan nilai modus setiap variabel dapat dilihat pada Tabel 4.



Tabel 4 Hasil Eksplorasi Data

Var	Data Terisi	Jumlah 0	Jumlah 8	Jumlah 9	Atribut Tidak Ada	Modus
R1	546	0	3	1	0	5
R2	546	0	3	0	9	5
R3	546	0	4	0	8-9	5
R4	546	0	0	0	9	5
R5	546	0	2	0	9	5
R6	546	0	1	0	8-9	5
R7	546	0	0	0	8-9	5
R8	546	0	0	0	4-5-6-7	5
R9	546	0	0	0	5	1
R10	546	0	0	0	-	1
R11	455	91	0	0	-	1
R12	546	0	0	0	-	72
R13	546	0	0	0	-	5
R14	546	0	0	0	7-8	3
R15	546	0	0	0	2-4-5-6-7	3
R16	546	0	0	0	7-9	5
R17	546	0	0	0	3	1
R18	526	20	0	0	-	1
R19	526	20	0	0	2-6	1
R20	546	0	0	0	11	2
R21	546	0	0	0	8	5
R22	546	0	0	0	1-11-2	4
R23	370	176	17	0	-	2
R24	546	0	0	0	4	1
R25	546	0	0	0	0-1-6-8-9-11	4
R26	546	0	0	0	-	5
R27	546	0	0	0	-	1
R28	546	0	0	0	-	5
R29	546	0	0	0	-	5
R30	546	0	0	0	-	5
R31	546	0	0	0	-	5
R32	546	0	0	0	-	5
R33	546	0	0	0	-	1
R34	546	0	0	0	-	5
R35	546	0	0	0	-	5
R36	546	0	0	0	-	5
R37	546	0	0	0	-	5
R38	546	0	0	0	-	1
R39	546	0	0	0	-	1
R40	35	511	0	0	-	2
R41	546	0	0	0	-	4

### 3.2 Data Preparation

Data preparation atau biasa disebut juga dengan pra-pemrosesan merupakan proses yang dilakukan untuk meningkatkan kualitas data dan meningkatkan efisiensi dan kemudahan penambangan data. Pada penelitian ini pra-pemrosesan yang dilakukan adalah dengan melakukan proses pembersihan data dan reduksi data.

#### 3.2.1 Pembersihan Data

Berdasarkan Tabel 4 dapat dilihat bahwa pada data yang digunakan masih ada data yang bernilai kosong atau tidak terdefinisi yaitu pada variabel R11, R18, R19, R23, dan R40, data bernilai 8 atau tidak tahu yaitu pada variabel R1, R2, R3, R5, R6, dan R23, serta data bernilai 9 atau



menolak menjawab pada variabel R23. Data kotor tersebut harus dibersihkan terlebih dahulu agar data dapat menghasilkan proses klasifikasi yang lebih baik. Proses yang dilakukan yaitu dengan mengubah data kosong tersebut dengan nilai modus dan atribut yang sesuai dengan kondisi pada data. Hasil pra-pemrosesan dapat dilihat pada Tabel 5.

Tabel 5 Hasil Pra-pemrosesan Pembersihan Data

Variabel	Total 0	Total 8	Total 9	Modus
R1	0	3	0	5
R2	0	3	0	5
R3	0	4	0	5
R5	0	2	0	5
R6	0	1	0	5
R11	91	0	0	1
R18	20	0	0	1
R19	20	0	0	1
R23	176	17	0	2
R40	511	0	0	2

Hasil pra-pemrosesan dengan teknik pembersihan data didapatkan hasil sebagai berikut:

- 1) Sebanyak 5 variabel yang memiliki data yang kosong yaitu, R11 dengan total data kosong sebanyak 91 data, R18 dengan total data kosong sebanyak 20, R19 dengan total data kosong sebanyak 20, R23 dengan total data kosong sebanyak 176, dan R40 dengan data kosong sebanyak 511.
- 2) Sebanyak 6 variabel yang memiliki data bernilai 8 dan 9 yaitu R1 dengan total data bernilai 8 sebanyak 3 data dan data bernilai 9 sebanyak 1 data, R2 dengan total data bernilai 8 sebanyak 3 data, R3 dengan total data bernilai 8 sebanyak 4 data, R5 dengan total data bernilai 8 sebanyak 2 data, R6 dengan total data bernilai 8 sebanyak 1 data, dan R23 dengan total data bernilai 8 sebanyak 17 data.

Data kotor tersebut diisi menggunakan nilai yang sering muncul atau modus masing-masing variabel dan atribut sesuai kondisi pada data.

### 3.2.2 Reduksi Data

Berdasarkan data pada Tabel 4 dapat dilihat bahwa pada data yang digunakan terdapat atribut yang tidak terpakai, yang mana atribut tersebut terdaftar pada meta data Susenas tetapi tidak ada pada data Susenas. Oleh karena itu, pada penelitian ini atribut yang tidak terpakai dihilangkan. Adapun atribut yang dihilangkan yaitu variabel R2, R3, R5, R6 dengan atribut 9, R4, R7, R8, dengan atribut 8 dan 9, R9 dengan atribut 4, 5, 6, dan 7, R10 dengan atribut 5, R14 dengan atribut 7 dan 9, R15 dengan atribut 2, 4, 5, 6, dan 7, R16 dengan atribut 7 dan 9, R17 dengan atribut 3, R19 dengan atribut 2 dan 6, R20 dengan atribut 11, R21 dengan atribut 8, R22 dengan atribut 1, 11, dan 2, R24 dengan atribut 4, dan R25 dengan atribut 0, 1, 6, 8, 9, dan 11.

### 3.3 Modeling

Pada tahap ini dilakukan proses pembuatan model dari metode yang digunakan yaitu metode Naïve Bayes. Pada proses pembuatan model klasifikasi menggunakan metode Naïve Bayes terlebih dahulu dilakukan proses *split* data untuk membagi data menjadi data latih dan data uji. Setelah itu dilakukan perhitungan probabilitas data latih kemudian dilanjutkan dengan melakukan perhitungan probabilitas terhadap data uji dan terakhir dilakukan pengujian terhadap data uji.

#### 3.3.1 Membagi Data

Proses pembagian data dilakukan dengan teknik *percentage split*. Pembagian data pada *percentage split* menggunakan 3 *scenario*, yang mana *scenario* 1 dengan jumlah perbandingan 60% (328 data latih)–40% (218 data uji), *scenario* 2 dengan jumlah perbandingan 70% (382 data



latih)–30% (164 data uji), dan *scenario* 3 dengan jumlah perbandingan 80% (437 data latihan)–20% (109 data uji).

### 3.3.2 Perhitungan Algoritma Naïve Bayes

#### 1) Perhitungan Probabilitas Kelas

Proses pertama dilakukan perhitungan probabilitas dari masing-masing kelas, yaitu probabilitas kelas layak dan probabilitas kelas tidak layak. Berikut adalah perhitungan probabilitas kelas:

$$S = 437$$

$$N_{(layak)} = 144$$

$$N_{(tidak layak)} = 293$$

$$P_{(layak)} = \frac{144}{437} = 0,3295$$

$$P_{(tidak layak)} = \frac{293}{437} = 0,6705$$

Di mana S merupakan jumlah seluruh data dengan N(Layak) dan N(Tidak Layak) adalah jumlah kelas “Layak” dan “Tidak Layak”. Sementara itu, P(Layak) dan P(Tidak Layak) merupakan jumlah probabilitas kelas “Layak” dan “Tidak Layak”.

#### 2) Perhitungan Probabilitas Atribut Tipe Data Kategorial

Langkah selanjutnya adalah menghitung probabilitas data uji. Atribut pada *dataset* memiliki tipe data kategorial dan numerik. Pada tahap ini dilakukan perhitungan nilai probabilitas data uji yang bertipe data kategorial menggunakan Pers. (2). Salah satu perhitungan probabilitas atribut dengan tipe data kategorial ada pada variabel “R26 - Apakah memiliki tabung gas 5,5 kg atau lebih?” yang memiliki atribut “1 = Ya”, dan “5=Tidak”. Contoh perhitungan probabilitas variabel tersebut adalah sebagai berikut:

$$P_{(1=Ya | 1=Layak)} = \frac{109}{144} = 0,7569444444444444$$

$$P_{(1=Ya | 0=Tidak Layak)} = \frac{234}{293} = 0,7986348122866894$$

$$P_{(5=Tidak | 1=Layak)} = \frac{35}{144} = 0,2430555555555556$$

$$P_{(5=Tidak | 0=Tidak Layak)} = \frac{59}{293} = 0,2013651877133106$$

#### 3) Perhitungan Probabilitas Atribut Numerik

Selain menghitung probabilitas data uji bertipe data kategorial selanjutnya juga dilakukan perhitungan probabilitas pada data uji yang bertipe data numerik dengan menggunakan Pers. (3). Salah satu perhitungan probabilitas atribut dengan tipe data numerik ada pada variabel “R12 - Berapa luas lantai rumah bangunan tempat tinggal?”. Contoh perhitungan probabilitas variabel tersebut dengan atribut bernilai 38 adalah sebagai berikut:



$$\mu = \frac{\sum_i^n x_i}{n}$$

$$\begin{aligned} \mu_{\text{kelas}="1"} &= \frac{(30 + 72 + 36 + 36 + 72 + 20 + \dots + 96 + 120 + 66 + 30 + 60 + 48 + 76)}{144} \\ &= \frac{9046}{144} = 62,81944 \end{aligned}$$

$$\begin{aligned} \mu_{\text{kelas}="0"} &= \frac{(81 + 54 + 63 + 112 + 40 + 40 + \dots + 90 + 50 + 80 + 32 + 60 + 72 + 60)}{293} \\ &= \frac{21480}{293} = 73,31058 \end{aligned}$$

$$\sigma = \sqrt{\frac{\sum_i^n (x_i - \mu)^2}{n - 1}}$$

$$\begin{aligned} \sigma_{\text{kelas}="1"} &= \sqrt{\frac{(30 - 62,81)^2 + (72 - 62,81)^2 + \dots + (48 - 62,81)^2 + (76 - 62,81)^2}{144 - 1}} \\ &= 27,47272 \end{aligned}$$

$$\begin{aligned} \sigma_{\text{kelas}="0"} &= \sqrt{\frac{(91 - 73,31)^2 + (54 - 73,31)^2 + \dots + (72 - 73,31)^2 + (60 - 73,31)^2}{293 - 1}} \\ &= 40,29552 \end{aligned}$$

$$g(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi} \cdot \sigma} \exp \frac{-(x - \mu)^2}{2\sigma^2}$$

$$g(8_1, \mu_1, \sigma_1) = \frac{1}{\sqrt{2} \times 3,14 \times 27,47272} \exp \frac{-(32 - 62,81944)^2}{2(27,47272)^2} = 0,0077$$

$$g(8_0, \mu_0, \sigma_0) = \frac{1}{\sqrt{2} \times 3,14 \times 40,29552} \exp \frac{-(32 - 73,31058)^2}{2(40,29552)^2} = 0,0058$$

### 3.4 Evaluasi Metode

Pengujian sistem penentuan masyarakat miskin penerima bantuan menggunakan metode *Confusion Matrix* untuk mendeteksi *accuracy*, *recall*, *precision*, dan *f1-score*. Proses pembagian data untuk mendapatkan hasil dari pengujian menggunakan *percentage split* dengan tiga *scenario*. Pengujian dilakukan berdasarkan keakuratan hasil klasifikasi data uji dengan data masyarakat yang layak dan tidak layak menerima bantuan sesuai dengan data Susenas yang telah didapatkan. Adapun *confusion matrix* hasil pengujian dari keseluruhan *scenario* pada *percentage split* dapat dilihat pada Tabel 6.

Tabel 6 Hasil *Confusion Matrix Percentage Split*

Keterangan	Persentase Data Latih dan Data Uji	TP	TN	FP	FN
<i>Scenario 1</i>	80%-20%	15	68	13	13
<i>Scenario 2</i>	70%-30%	24	99	19	22
<i>Scenario 3</i>	60%-40%	43	127	22	26

Hasil pengujian perolehan *accuracy*, *recall*, *precision*, dan *f1-score* seluruh data uji dapat dilihat pada Tabel 7.



Tabel 7 Hasil Pengujian *Percentage Split*

Keterangan	Persentase Data Latih dan Data Uji	Accuracy	Recall	Precision	F1-Score
Scenario 1	80%-20%	76,15%	53,57%	53,57%	53,57%
Scenario 2	70%-30%	75%	52,17%	55,80%	53,93%
Scenario 3	60%-40%	77,98%	62,32%	66,15%	64,18%

Berdasarkan Tabel 7, dapat disimpulkan bahwa sistem penentuan kelayakan masyarakat miskin penerima bantuan yang dibuat dapat melakukan prediksi dengan cukup baik pada *scenario* ketiga yaitu 60% data latih (328 data) dan 40% data uji (218 data). Hasil *accuracy* yang didapatkan adalah 77,98% dengan *recall* 62,32%, *precision* 66,15%, dan *F1-score* 64,18%.

Pada penelitian ini dilakukan perbandingan hasil yang diperoleh dari metode, variabel, dan jumlah data yang diusulkan dengan metode, variabel, dan jumlah data yang digunakan penelitian lain. Hal ini dilakukan untuk mengevaluasi model yang dibangun. Adapun perbandingan penelitian sebelumnya ditunjukkan pada Tabel 8.

Tabel 8 Perbandingan Kinerja Metode yang diusulkan dengan Metode Lain

No.	Sitasi	Parameter	Klasifikasi	Jumlah Data	Akurasi
1	Annur (2018)	Pendidikan, Pekerjaan, Penghasilan, Tanggungan, dan Status Perkawinan.	Naïve Bayes	190	73%
2	Aji (2019)	Luas Bangunan, Jenis Lantai, Jenis Dinding, Fasilitas MCK, Sumber Air Minum, Sumber Penerangan, Pekerjaan Kepala Keluarga, Penghasilan, Kondisi Rumah, Jumlah Tanggungan, Bahan Bakar Memasak, dan Kepemilikan Asset (status Kepemilikan Rumah).	Naïve Bayes	300	80%
3	Arifando et al. (2017)	Jumlah Anggota Keluarga, Pendapatan, Umur, Kondisi Rumah, Status Kepemilikan Rumah, Pengeluaran, dan Pendidikan Terakhir.	<i>Learning Vector Quantization</i> (LVQ)	N/A	98%
4	Fitriani (2020)	Lansia, Pendidikan Terakhir, Anak Sekolah, Pekerjaan, Ibu Hamil, Membeli Pakaian, Frekuensi Makanan, Berobat Ke Puskesmas, Membeli (Daging Ayam, Susu), Aset, Jenis Dinding, Jenis Lantai, Sumber Penerangan, Sumber Air Minum, Jenis Bahan Bakar Memasak, Fasilitas BAB, dan Luas Lantai.	C4.5 Naïve Bayes	1.109	91,25% 87,11%
5	Purnama et al. (2020)	Jumlah Tanggungan, Kondisi Rumah, Sumber Air Minum, Ketersediaan WC, dan Bahan Bakar.	Naïve Bayes	210	82,14%
6	Kurnia et al. (2019)	Kartu Keluarga, Jumlah Anggota Keluarga, Jenis Pekerjaan, dan Penghasilan Bulanan.	K- Nearest Neighbor	100	90%



Tabel 9 Perbandingan Kinerja Metode yang diusulkan dengan Metode Lain (Lanjutan)

No.	Sitasi	Parameter	Klasifikasi	Jumlah Data	Akurasi
7	Nurmayanti et al. (2021)	Jenis Kelamin, Peserta PKH, Jumlah Anggota Rumah Tangga, Status Kepemilikan Rumah, Fasilitas Rumah, Luas Lantai Rumah, dan Fasilitas Elektronik.	Naïve Bayes	168	96,63%
8	Ramadani et al. (2020)	Jumlah Tanggungan, Kepala Rumah Tangga, Status Tempat Tinggal, Jenis Lantai, Jenis Dinding, Pendapatan, dan Sumber Penerangan.	Naïve Bayes	20	-
9	Riyanah & Fatmawati (2021)	Pekerjaan, Penghasilan, Usia, Status, Kendaraan, Kepemilikan Rumah, dan Atap Bangunan.	Naïve Bayes	35	62,86%

Berdasarkan Tabel 8 dapat diketahui bahwa terdapat penelitian yang memiliki variabel dan jumlah data lebih sedikit dari penelitian ini. Hasil penelitian yang dilakukan oleh Annur (2018) dan Riyanah & Fatmawati (2021) memiliki tingkat akurasi yang lebih kecil dari penelitian yang dilakukan, sedangkan penelitian yang lainnya menghasilkan tingkat akurasi yang lebih besar dari penelitian yang dilakukan.

#### 4. KESIMPULAN

Berdasarkan implementasi dan hasil pengujian pada penelitian penentuan kelayakan masyarakat miskin penerima bantuan dapat diambil beberapa kesimpulan yaitu metode Naïve Bayes dapat diterapkan pada sistem penentuan kelayakan masyarakat miskin penerima bantuan karena mendapatkan *accuracy* yang cukup baik, dan berdasarkan hasil pengujian menunjukkan bahwa metode Naïve Bayes dapat menghasilkan *accuracy* tertinggi pada scenario ketiga yaitu 60% atau 328 data latihan dan 40% atau 218 data uji dengan nilai *accuracy* sebanyak 77,98%. Pada penelitian lebih lanjut dapat dikembangkan dengan menggunakan metode lain seperti *Artificial Neural Network*, *Support Vector Machine*, atau *K-Nearest Neighbor* untuk meningkatkan nilai akurasi.

#### DAFTAR PUSTAKA

- Aji, A. (2019). *Penerapan Metode Naive Bayes untuk Mengklasifikasi Kelayakan Penerima Bantuan Beras Miskin (Studi Kasus: Kantor Kelurahan Desa Tegalyoso)* [Universitas Teknologi Yogyakarta]. <http://eprints.uty.ac.id/2660/>
- Annur, H. (2018). Klasifikasi Masyarakat Miskin Menggunakan Metode Naive Bayes. *ILKOM Jurnal Ilmiah*, 10(2), 160–165. <https://doi.org/10.33096/ilkom.v10i2.303.160-165>
- Arifando, R., Hidayat, N., & Soebroto, A. A. (2017). Klasifikasi Calon Penerima Bantuan Keluarga Miskin Menggunakan Metode Learning Vector Quantization (LVQ) (Studi Kasus: Daerah Kecamatan Mlandingan, Situbondo). *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 2(6), 2173–2181. <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/1625>
- Badan Pusat Statistik. (2021). *Profil Kemiskinan di Penajam Paser Utara Tahun 2021*. Badan Pusat Statistik. <https://ppukab.bps.go.id/pressrelease/2021/12/22/229/profil-kemiskinan-di-kabupaten-penajam-paser-utara-tahun-2021.html>
- Fitriani, E. (2020). Perbandingan Algoritma C4.5 dan Naïve Bayes untuk Menentukan Kelayakan Penerima Bantuan Program Keluarga Harapan. *SISTEMASI*, 9(1), 103. <https://doi.org/10.32520/stmsi.v9i1.596>
- Gunawan, B., Pratiwi, H. S., & Pratama, E. E. (2018). Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode Naive Bayes. *Jurnal Edukasi Dan Penelitian Informatika (JEPIN)*, 4(2), 113. <https://doi.org/10.26418/jp.v4i2.27526>
- Kurnia, F., Kurniawan, J., Fahmi, I., & Monalisa, S. (2019). Klasifikasi Keluarga Miskin Menggunakan Metode K-Nearest Neighbor Berbasis Euclidean Distance. *Seminar Nasional*



- Teknologi Informasi Komunikasi Dan Industri*, 230–239. <https://ejournal.uin-suska.ac.id/index.php/SNTIKI/article/view/8089>
- Kurniawan, D. (2020). *Pengenalan Machine Learning dengan Python*. PT Elex Media Komputindo.
- Maarif, M. R. (2016). Perbandingan Naïve Bayes Classifier dan Support Vector Machine untuk Klasifikasi Judul Artikel. *JISKA (Jurnal Informatika Sunan Kalijaga)*, 1(2), 90–93. <https://doi.org/10.14421/jiska.2016.12-05>
- Muljono, Artanti, D. P., Syukur, A., Prihandono, A., & Setiadi, D. R. I. M. (2018). *Analisa Sentimen Untuk Penilaian Pelayanan Situs Belanja Online Menggunakan Algoritma Naïve Bayes*. 8–9.
- Nastuti, A., & Harahap, S. Z. (2019). Teknik Data Mining untuk Penentuan Paket Hemat Sembako dan Kebutuhan Harian Dengan Menggunakan Algoritma FP-Growth (Studi Kasus di Ulfamart Lubuk Alung). *Jurnal Informatika*, 7(3), 111–119. <https://doi.org/10.36987/informatika.v7i3.1381>
- Natuzzuhriyyah, A., Nafisah, N., & Mayasari, R. (2021). Klasifikasi Tingkat Kepuasan Mahasiswa Terhadap Pembelajaran Secara Daring Menggunakan Algoritma Naïve Bayes. *JISKA (Jurnal Informatika Sunan Kalijaga)*, 6(3), 161–170. <https://doi.org/10.14421/jiska.2021.6.3.161-170>
- Novriansyah, M. A. (2018). Pengaruh Pengangguran dan Kemiskinan Terhadap Pertumbuhan Ekonomi di Provinsi Gorontalo. *Gorontalo Development Review*, 1(1), 59–73. <https://doi.org/10.32662/GOLDER.V1i1.115>
- Nurmawanti, W. P., Saky, D. A. L., Malthuf, M., Gazali, M., & Hirzi, R. H. (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>
- Purnama, A. I., Aziz, A., & Sartika Wiguna, A. (2020). Penerapan Data Mining untuk Mengklasifikasi Penerima Bantuan PKH Desa Wae Jare Menggunakan Metode Naïve Bayes. *KURAWAL Jurnal Teknologi, Informasi Dan Industri*, 3(2), 173–180. <https://jurnal.machung.ac.id/index.php/kurawal>
- Putri, A. C., Hariyanto, F. E., Andini, N. L. E., & Zulkarnaen, Z. C. S. (2021). Klasifikasi Rumah Tangga Miskin di Provinsi Papua Tahun 2017 Menggunakan Metode Naïve Bayes. *Jurnal Sains Matematika Dan Statistika*, 7(1), 89. <https://doi.org/10.24014/jsms.v7i1.11924>
- Ramadani, S., Zannah, N., Ayu, S., Nurhayati, N., Azzahra, F., & Windarto, A. P. (2020). Analisis Data Mining Naive Bayes Klasifikasi Pada Kelayakan Penerima PKH. *KOMIK (Konferensi Nasional Teknologi Informasi Dan Komputer)*, 4(1). <https://doi.org/10.30865/KOMIK.V4i1.2726>
- Riyannah, N., & Fatmawati, F. (2021). Penerapan Algoritma Naive Bayes Untuk Klasifikasi Penerima Bantuan Surat Keterangan Tidak Mampu. *JTIM : Jurnal Teknologi Informasi Dan Multimedia*, 2(4), 206–213. <https://doi.org/10.35746/jtim.v2i4.117>
- Setiawan, D. A., Halilintar, R., & Wahyuniar, L. S. (2021). Penerapan Metode Naive Bayes Untuk Klasifikasi Penentuan Penerima Bantuan PKH. *Prosiding SEMNAS INOTEK (Seminar Nasional Inovasi Teknologi)*, 5(2), 249–254. <https://proceeding.unpkediri.ac.id/index.php/inotek/article/view/1137>
- Sugianto, C. A., & Maulana, F. R. (2019). Algoritma Naïve Bayes Untuk Klasifikasi Penerima Bantuan Pangan Non Tunai ( Studi Kasus Kelurahan Utama ). *Techno.Com*, 18(4), 321–331. <https://doi.org/10.33633/tc.v18i4.2587>
- Suntoro, J. (2019). *Data Mining Algoritma dan Implementasi dengan Pemrograman PHP*. PT Elex Media Komputindo. <https://elexmedia.id/produk/detail/elexmedia2018-data-mining-algoritma-dan-implementasi-dengan-pemrograman-php/9786020498812>
- Umami, U. (2013). Cara Pandang dan Upaya Pemerintah dalam Mengurangi Kemiskinan. *Jurnal Pembangunan Wilayah & Kota*, 9(4), 343. <https://doi.org/10.14710/pwk.v9i4.6673>
- Wati, M., & Hadi, A. (2016). Implementasi Algoritma Naive Bayesian Dalam Penentuan Penerima Program Bantuan Pemerintah. *JTRISTE*, 3(1), 22–26.

