

Klasifikasi Tingkat Rumah Tangga Miskin Saat Pandemi Dengan Naïve Bayes Classifier

Harliana¹⁾, Fatra Nonggala Putra²⁾

¹⁾²⁾ Program Studi Ilmu Komputer, Fakultas Eksakta, Universitas Nahdlatul Ulama Blitar
Jl Masjid No. 22, Kota Blitar

¹⁾ harliana@unublitar.ac.id

²⁾ fatranp@unublitar.ac.id

Abstrak

Secara definisi kemiskinan merupakan suatu kondisi individu ditingkat rumah tangga yang dinilai berdasarkan karakteristik kemiskinan. Sebagai dampak dari pandemi covid-19 persentase rumah tangga miskin di Indonesia meningkat sekitar 9,78%. Berdasarkan hal tersebut, maka penelitian ini akan melakukan klasifikasi dengan algoritma Naïve Bayes Classification untuk menentukan rumah tangga miskin melalui parameter survey ekonomi Nasional Tahun 2020 Modul Ketahanan Sosial yang berfokus pada pengeluaran dan konsumsi perkapita responden selama pandemi. Sedangkan tujuan dari penelitian ini adalah mendapatkan akurasi tertinggi yang dihasilkan oleh *Naïve Bayes Classifier* dalam penentuan rumah tangga miskin. Menurut hasil pengujian dengan *confusion matrix* dan *10-fold cross validation* didapatkan bahwa rata-rata akurasi tertinggi terjadi pada fold ke-10 dengan nilai *accuracy* 93,21%; *precision* 86,3%; dan *recall* 80,11%. Hal ini berarti bahwa akurasi yang dihasilkan oleh *naïve bayes classifier* dalam melakukan klasifikasi rumah tangga miskin cukup tinggi.

Kata kunci: kemiskinan, *naïve bayes classification*, akurasi, *confusion matrix*, *k-fold cross validation*.

Abstract

By definition, poverty is an individual condition at the household level which is assessed based on the characteristics of poverty. As a result of the COVID-19 pandemic, the percentage of poor households in Indonesia increased by around 9.78%. Based on this, this study will classify using the Naïve Bayes Classification algorithm to determine poor households through the parameters of the 2020 National Economic Survey, the Social Security Module, which focuses on the respondents' expenditure and consumption per capita during the pandemic. Meanwhile, the purpose of this study is to obtain the highest accuracy produced by the Naïve Bayes Classification in determining poor households. According to the test results with the confusion matrix and 10-fold cross-validation, it was found that the highest average accuracy occurred in the 10th fold with an accuracy value of 93.21%; precision of 86.3%; and 80.11% recalls. This means that the accuracy produced by the Naïve Bayes classifier in classifying poor households is quite high.

Keywords: *poverty, naïve bayes classification, accuracy, confusion matrix, k-fold cross validation.*

1. PENDAHULUAN

Kemiskinan merupakan salah satu permasalahan yang masih menjadi pokok utama pemerintah. Dalam menentukan suatu penduduk tergolong miskin atau tidak saat ini pemerintah masih menggunakan sampel rumah tangga miskin, dimana secara definisi rumah tangga miskin adalah suatu kondisi individu yang akan dinilai berdasarkan karakteristik kemiskinan yang telah ditetapkan [1]. Perbulan Maret 2020 Badan Pusat Statistik Jawa Barat mencatat jumlah penduduk miskin di Jawa Barat mengalami kenaikan sekitar 6,28% [2]. Sedangkan menurut data BPS Nasional, tingkat kemiskinan di Indonesia diperkirakan

akan meningkat menjadi 9,78% atau naik sekitar 0,37% sebagai dampak dari pandemi Covid-19 yang mengakibatkan konsumsi rumah tangga akan melambat. Untuk menangani kemiskinan tersebut, pemerintah telah mengeluarkan berbagai bantuan untuk masing-masing *cluster* kemiskinan menjadi *cluster 1*, *cluster 2*, dan *cluster 3*. *Cluster 1* merupakan bantuan untuk kelompok Rumah Tangga Sangat Miskin (RTSM) dimana pemerintah akan memberikan bantuan sosial terpadu berbasis keluarga, tujuannya adalah mengurangi beban rumah tangga miskin melalui peningkatan akses pelayanan kesehatan, pendidikan, air bersih dan sanitasi [3]. Pada *cluster 2*, pemerintah akan memberikan bantuan berupa pemberdayaan masyarakat dengan tujuan mengembangkan potensi dan memperkuat kapasitas kelompok masyarakat miskin [3]. Sedangkan pada *cluster 3*, pemerintah akan melakukan penanggulangan kemiskinan berbasis pemberdayaan usaha ekonomi mikro dan kecil dengan tujuan memberikan akses dan penguatan ekonomi bagi pelaku usaha [3].

Pada masa pandemi saat ini, segala jenis bantuan pemerintah serasa tidak dapat merubah tingkat kemiskinan yang terjadi, bahkan ada sebagian penduduk yang justru masuk ke dalam *cluster 1* yang sebelumnya berada di *cluster 2 / 3*. Penduduk yang berada pada *cluster 1* merupakan kelompok informal yang paling rentan mengalami kemiskinan *cyclical poverty* akibat pola siklus ekonomi secara keseluruhan [4]. Segala macam perubahan kelompok ini dapat disebabkan oleh berubahnya pola konsumsi pangan, serta menurunnya pendapatan masyarakat pada saat *work from home (wfh)* dan Pemberlakuan Pembatasan Kegiatan Masyarakat (PPKM) yang diterapkan pada beberapa wilayah. Berdasarkan hal tersebut, maka penelitian ini akan melakukan klasifikasi terhadap rumah tangga miskin baru berdasarkan kebutuhan dasarnya. Adapun tujuan dari penelitian ini yaitu melihat akurasi yang dihasilkan oleh *Naïve Bayes Classification* terhadap penentuan tingkat kesejahteraan Rumah Tangga Miskin di masa Pandemi Covid-19. Hasil dari penelitian ini diharapkan dapat mendapatkan status rumah tangga miskin yang lebih akurat apakah masuk kedalam RTSM (Rumah Tangga Sangat Miskin), RTM (Rumah Tangga Miskin) ataukah RTTM (Rumah Tangga Tidak Miskin) dimana dalam melakukan penentuan ini tidak lagi menggunakan pengelompokan secara bersama tetapi melalui suatu klasifikasi. Dalam penelitian ini akan menggunakan *Naïve bayes classification* dalam melakukan klasifikasinya, *Naïve Bayes Classification* dipilih karena kemampuannya dalam menangani data yang besar, penanganan *missing value* terhadap dataset, serta jumlah atribut yang tidak sama [5]. Untuk data yang sederhana *Naïve bayes* mampu menghasilkan perhitungan yang cepat dengan tingkat akurasi yang tinggi, namun apabila dibandingkan dengan C4.5 akurasi yang dihasilkan oleh *Naïve Bayes* ternyata lebih kecil [6].

Dalam menentukan penduduk miskin beberapa parameter yang biasa digunakan diantaranya umur, tanggungan, pekerjaan, pendidikan, status [7], pendapatan perbulan, keikutsertaan program KPS, KIS/JKN, PKH [8], jenis kelamin, lapangan usaha, jenis atap terluas, jenis lantai terluas, jenis dinding terluas, sumber air minum, sumber penerangan, dan bahan bakar untuk memasak [9]. Namun dalam penelitian ini penentuan penduduk miskin akan menggunakan parameter berdasarkan data Survey Ekonomi Nasional Tahun 2020 Modul Ketahanan Sosial dengan fokus pengeluaran dan konsumsi perkapita responden selama Pandemi.

2. TINJAUAN PUSTAKA

Beberapa penelitian terdahulu tentang penerapan *naïve bayes* pernah dilakukan untuk mengklasifikasi penerima bantuan sosial PKH di Kecamatan Cimanggu, dari 17 atribut yang digunakan hanya 13 atribut yang memiliki pengaruh besar terhadap penentuan penerima PKH diantaranya jumlah anggota rumah tangga yang lebih dari 3 orang, tidak memiliki lemari es, telepon, laptop, perahu motor, kapal dan berbagai asset tak bergerak lainnya [10]. Selain di Kecamatan Cimanggu, *naïve bayes* juga pernah digunakan untuk menentukan penerima PKH di Desa Minggiran Kediri dengan hanya menggunakan 5 atribut yaitu status kepala keluarga, jumlah tanggungan keluarga, usia kepala keluarga, dan penghasilan kepala keluarga. Meskipun hanya menggunakan 5 atribut namun penelitian ini mampu menghasilkan akurasi sebesar 93,33% dalam menentukan calon penerima PKH dengan menggunakan *naïve bayes* [11].

Apabila dilihat dari akurasinya, rata-rata *naïve bayes* memiliki nilai akurasi lebih baik bila dibandingkan dengan algoritma klasifikasi yang lain, hal ini terlihat dari perbandingan antara

naïve bayes dengan k-nearest neighbor dalam memprediksi status pekerjaan alumni. Berdasarkan hasil analisa maka akurasi dan MAPE yang dihasilkan oleh naïve bayes masih sedikit lebih baik bila dibandingkan dengan k-nearest neighbor yaitu 83,83% untuk akurasinya dan 16,17% untuk MAPE yang dihasilkan [12]. Namun apabila dibandingkan dengan *Support Vector Machine* (SVM) dalam mendeteksi serangan pada jaringan computer, akurasi yang dihasilkan oleh naïve bayes cenderung lebih rendah dengan nilai AUC pada kurva ROC yang hanya bernilai 0,50 dengan garis kurva sama persis di garis *threshold* [13].

Melalui kelebihan-kelebihan dari naïve bayes tersebut, maka pada penelitian ini peneliti memutuskan untuk menggunakan naïve bayes dalam melakukan klasifikasi terhadap tingkat penduduk miskin. Adapun alur dari metode ini adalah:

- a. Baca data training
- b. Hitung jumlah dari probabilitas, dimana:
 - 1) Jika data adalah numerik, maka cari nilai mean dan standar deviasi dari masing-masing parameter yang menggambarkan data angka. Persamaan (1) dapat digunakan untuk menghitung mean tersebut.

$$\mu = \sum_{i=1}^n x_i \quad (1)$$

Atau bisa juga menggunakan persamaan (2).

$$\mu = \frac{x_1 + x_2 + \dots + x_n}{n} \quad (2)$$

Dimana:

- μ : rata-rata hitung (mean)
- x_i : nilai sample ke-i
- n : jumlah sample

dan persamaan untuk menghitung simpangan baku (standar deviasi) terdapat pada persamaan (3)

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \mu)^2}{n - 1}} \quad (3)$$

Dengan:

- σ : standar deviasi
- x_i : nilai x ke i
- μ : rata-rata hitung
- n : jumlah sample

- 2) Jika data bukan numerik, maka hitung nilai probabilitas tiap kategori yang sama dengan cara menjumlahkan data yang sesuai dari kategori yang sama lalu dibagi dengan jumlah data pada kategori tersebut dalam menentukan probabilitiknya.

3. METODE PENELITIAN

Dalam penelitian ini akan digunakan 368 dataset kemiskinan di Kecamatan Gebang Kabupaten Cirebon dengan 21 parameter. Sedangkan metode penelitiannya menggunakan CRISP-DM (*Cross-Standard Industry for Data Mining*) dengan 6 tahapan, yaitu:

a. *Business understanding*

Pada tahapan ini akan dilakukan identifikasi terhadap permasalahan yang terjadi dalam menentukan status Rumah Tangga miskin berdasarkan kebutuhan dasarnya menurut hasil Survey Sosial Ekonomi Nasional Tahun 2020 modul ketahanan pangan. Pada tahapan ini juga ditentukan tujuan dari penelitian yaitu melihat akurasi dari *Naïve Bayes Classifier* dalam melakukan klasifikasi rumah tangga miskin.

b. *Data understanding*

Pada tahapan ini akan dilakukan pengumpulan data, mulai dari pemahaman atas data, analisis terhadap data serta parameter yang akan digunakan. Penelitian ini menggunakan data kemiskinan Kecamatan Gebang Kabupaten Cirebon Tahun 2020. Dimana dataset yang akan digunakan berjumlah 368 data dengan 20 atribut dan 1 label. Tabel 1 adalah rangkuman dari atribut dan label yang akan digunakan.

Tabel 1. Atribut dan label yang akan digunakan

No	Data input	Jenis
1	Hubungan dengan kepala rumah tangga	Atribut
2	Status perkawinan	Atribut
3	Jenis kelamin	Atribut
4	Umur	Atribut
5	Sarana angkutan yang digunakan	Atribut
6	Pendidikan terakhir	Atribut
7	Jenis pekerjaan	Atribut
8	Kepemilikan kartu prakerja	Atribut
9	Kepemilikan tempat tinggal	Atribut
10	Luas lantai bangunan	Atribut
11	Akses listrik	Atribut
12	Kepemilikan jaminan sosial	Atribut
13	Kepemilikan Kartu Keluarga Sejahtera (KKS)	Atribut
14	Penerima PKH (Program Keluarga Harapan)	Atribut
15	Penerima BPNT (Bantuan Pangan Non Tunai)	Atribut
16	Penerima BST (Bantuan Sosial Tunai)	Atribut
17	Penerima BLT (Bantuan Langsung Tunai)	Atribut
18	Prosentase kenaikan/penurunan total pengeluaran makanan	Atribut
19	Prosentase kenaikan/penurunan total pengeluaran non-makanan	Atribut
20	Prosentase kenaikan/penurunan total pengeluaran keseluruhan	Atribut
21	Status	Label

c. *Data preparation*

Pada tahapan ini, akan dilakukan penanganan terhadap *missing values* di dataset. Dari 368 dataset kemiskinan yang didapatkan terdapat 20 data yang mengandung *missing values*. Karena dataset yang digunakan bersifat numerik maka *missing value* ditangani dengan mencari nilai *mean* dari keseluruhan data.

d. *Modelling*

Pada tahapan ini akan dilakukan pemodelan dataset dengan menggunakan *Naïve Bayes Classifier* untuk mengetahui akurasi yang dihasilkan dalam mengklasifikasikan status rumah tangga miskin.

e. *Evaluation*

Pada tahapan ini akan dilakukan evaluasi akurasi yang dihasilkan oleh *Naïve Bayes Classifier* melalui *confusion matrix*, sedangkan untuk mengetahui kinerja yang dihasilkan akan diuji dengan menggunakan *k-fold cross validation* sebanyak 10 kali (*10-fold cross validation*)

f. *Deployment*

Pada tahapan ini akan didapatkan akurasi terhadap hasil klasifikasi rumah tangga miskin.

4. PEMBAHASAN

Penjelasan dari masing-masing atribut yang digunakan adalah:

- a. Hubungan dengan kepala keluarga. Atribut ini akan dikelompokkan menjadi: kepala rumah tangga dengan nilai 1, istri / suami dengan nilai 2, anak kandung / tiri dengan nilai 3, anak angkat / tiri dengan nilai 4, menantu dengan nilai 5.
- b. Status perkawinan. Atribut ini akan dikelompokkan menjadi: belum kawin dengan nilai 1, kawin dengan nilai 2, cerai hidup dengan nilai 3, dan cerai mati dengan nilai 4.
- c. Jenis kelamin. atribut ini akan dikelompokkan menjadi: laki-laki dengan nilai 1, dan perempuan dengan nilai 2.
- d. Umur, merupakan atribut yang digunakan untuk mengetahui usia dari responden.
- e. Sarana angkutan yang digunakan, merupakan atribut yang digunakan untuk mengetahui keseharian responden dalam beraktifitas. Atribut ini akan dikelompokkan menjadi: tidak menggunakan kendaraan dengan nilai 1, kendaraan tidak bermotor dengan nilai 2, kendaraan bermotor umum dengan rute tertentu dengan nilai 3, kendaraan bermotor umum tidak berute dengan nilai 4, dan kendaraan bermotor pribadi dengan nilai 5.
- f. Pendidikan terakhir. Atribut ini akan dikelompokkan menjadi: tidak tamat SD dengan nilai 1, SD dengan nilai 2, SMP dengan nilai 3, SMA dengan nilai 4, Perguruan Tinggi / Universitas dengan nilai 5.
- g. Jenis pekerjaan. Atribut ini akan dikelompokkan menjadi: pertanian, tanaman padi dan palawija dengan nilai 1, hortikultura dengan nilai 2, perkebunan dengan nilai 3, perikanan dengan nilai 4, peternakan dengan nilai 5, kehutanan dan lainnya dengan nilai 6, pertambangan dan penggalian dengan nilai 7, industry pengolahan dengan nilai 8, pengadaan listrik dan sejenisnya dengan nilai 9, pengelolaan air dan sejenisnya dengan nilai 10, kontruksi dengan nilai 11, perdagangan besar dan eceran, reparasi dan perawatan mobil/motor dengan nilai 12, pengangkutan dan pergudangan dengan nilai 13, penyediaan akomodasi dan penyediaan makan minum dengan nilai 14, infomasi dan komunikasi dengan nilai 15, aktivitas keuangan dengan nilai 16, real estate dengan nilai 17, aktivitas professional, ilmiah dan teknis dengan nilai 18, aktivitas penyewaan dan sewa guna dan sejenisnya dengan nilai 19, administrasi pemerintahan dengan nilai 20, pendidikan dengan nilai 21, aktivitas kesehatan manusia dan sosial dengan nilai 22, kesenian, hiburan dan rekreasi dengan nilai 23, aktivitas jasa lainnya 24, aktivitas rumah tangga sebagai pemberi kerja dengan nilai 25, aktivitas badan internasional dan sejenisnya dengan nilai 26.
- h. Kepemilikan kartu prakerja. Atribut ini akan dikelompokkan menjadi jika Ya akan diberi nilai 1, dan jika Tidak maka akan diberi nilai 5.
- i. Kepemilikan tempat tinggal. Atribut ini akan dikelompokkan menjadi: milik sendiri dengan nilai 1, kontrak/sewa dengan nilai 2, bebas sewa dengan nilai 3, dinas dengan nilai 4.
- j. Akses listrik. Atribut ini akan dikelompokkan menjadi: PLN dengan daya 1300watt atau lebih dengan nilai 1, PLN dengan daya antara 450watt – 900watt dengan nilai 2, PLN tanpa meteran 3, listrik non-PLN dengan nilai 4, tidak menggunakan listrik dengan nilai 5.
- k. Kepemilikan jaminan sosial. Atribut ini akan dikelompokkan menjadi 2, yaitu jika Ya memiliki jaminan sosial akan diberi nilai 1 dan jika Tidak maka akan diberi nilai 2.
- l. Kepemilikan KKS. Atribut ini akan dikelompokkan menjadi 3, yaitu jika Ya dan dapat menunjukkan kartu maka akan diberi nilai 1, jika Ya tetapi tidak dapat menunjukkan kartu maka diberi nilai 2, dan jika Tidak maka diberikan nilai 3.
- m. Untuk kepemilikan PKH, BPNT, BST, BLT akan dikelompokkan dengan cara yang sama yaitu jika Ya maka akan diberikan nilai 1 dan jika Tidak akan diberikan nilai 2.
- n. Untuk prosentase kenaikan / penurunan total pengeluaran makan, non-makan, dan total keseluruhan akan dikelompokkan dengan cara yang sama juga yaitu jika tetap / sama akan diberikan nilai 1, naik diatas 10% diberikan nilai 2, naik antara 5% - 10% diberikan nilai 3, naik kurang dari 5% diberikan nilai 4, turun diatas 10% diberikan nilai 5, turun antara 5% - 10% diberikan nilai 6, turun kurang dari 5% diberikan nilai 7.

Karena data yang digunakan bersifat numerik, maka akan dicari nilai *mean* dan standar deviasi terlebih dahulu. Nilai *mean* dihitung berdasarkan hasil pembagian dari keseluruhan data dengan banyaknya data[14]. Sedangkan standar deviasi digunakan untuk menentukan sebaran data dari *sample*[14]. Pada penelitian ini akan digunakan 110 data training (30% data) dengan jumlah RTSM (C_0) sebanyak 54 data, RTM (C_1) sebanyak 33 data, dan RTTM (C_2) sebanyak 23 data. Dari data tersebut, maka nilai dari probabilitas prior dari ketiganya adalah:

$$P(C_0) = \frac{54}{110} = 0,49$$

$$P(C_1) = \frac{33}{110} = 0,3$$

$$P(C_2) = \frac{23}{110} = 0,209$$

Langkah selanjutnya yaitu mencari nilai *probabilitas posterior* dengan masing-masing atribut X yang digunakan. Tabel 2 adalah hasil *probabilitas* untuk pendidikan terakhir responden.

Tabel 2. Probabilitas pendidikan terakhir

Jenis pendidikan	Jumlah kejadian dipilih			Probabilitas		
	RTSM	RTM	RTTM	RTSM	RTM	RTTM
Tidak tamat	12	6	2	0,1578	0,2727	0,1666
SD	19	2	1	0,2500	0,0909	0,0833
SMP	27	4	3	0,3552	0,1818	0,2500
SMA	18	10	4	0,2368	0,4545	0,3333
PT /	0	0	2	0	0	0,1666
Jumlah	76	22	12	1	1	1

Untuk nilai *probabilitas* kenaikan / penurunan pengeluaran makan terangkum pada Tabel 3.

Tabel 3. Probabilitas kenaikan / penurunan pengeluaran makan

Jenis kenaikan / penurunan	Jumlah kejadian dipilih			Probabilitas		
	RTSM	RTM	RTTM	RTSM	RTM	RTTM
tetap / sama	2	0	2	0,0294	0	0,2500
Naik diatas 10%	10	2	1	0,1470	0,0740	0,1250
Naik diantara 5% - 10%	24	3	1	0,3529	0,1111	0,1250
Naik kurang dari 5% - 10%	32	22	2	0,4705	0,8148	0,2500
Turun diatas 10%	0	0	0	0	0	0
Turun diantara 5% - 10%	0	0	1	0	0	0,1250
Turun kurang dari 5%	0	1	1	0	0,0370	0,1250
Jumlah	68	27	8	1	1	1

Untuk nilai *probabilitas* kenaikan / penurunan pengeluaran makan terangkum pada Tabel 4.

Tabel 4. Probabilitas kenaikan / penurunan pengeluaran non-makan

Jenis kenaikan / penurunan	Jumlah kejadian dipilih			Probabilitas		
	RTSM	RTM	RTTM	RTSM	RTM	RTTM
tetap / sama	10	1	10	0,4545	0,03125	0,7154
Naik diatas 10%	1	0	3	0,0454	0	0,0526
Naik diantara 5% - 10%	0	2	7	0	0,0625	0,1228
Naik kurang dari 5% - 10%	3	9	5	0,1363	0,28125	0,0877
Turun diatas 10%	0	0	2	0	0	0,0350
Turun diantara 5% - 10%	6	11	11	0,2727	0,34375	0,1929
Turun kurang dari 5%	2	9	19	0,0909	0,28125	0,3333
Jumlah	22	32	57	1	1	1

Sedangkan untuk nilai *probabilitas* kenaikan / penurunan total pengeluaran keseluruhan terangkum pada Tabel 5.

Tabel 5. Probabilitas kenaikan / penurunan total pengeluaran keseluruhan

Jenis kenaikan / penurunan	Jumlah kejadian dipilih			Probabilitas		
	RTSM	RTM	RTTM	RTSM	RTM	RTTM
tetap / sama	20	26	9	0,2325	0,6341	0,4285
Naik diatas 10%	25	6	5	0,2906	0,1463	0,2380
Naik diantara 5% - 10%	21	7	4	0,2441	0,1707	0,1904
Naik kurang dari 5% - 10%	17	2	2	0,1976	0,0487	0,0952
Turun diatas 10%	0	0	0	0	0	0
Turun diantara 5% - 10%	1	0	1	0,0116	0	0,0476
Turun kurang dari 5%	2	0	0	0,0232	0	0
Jumlah	86	41	21	1	1	1

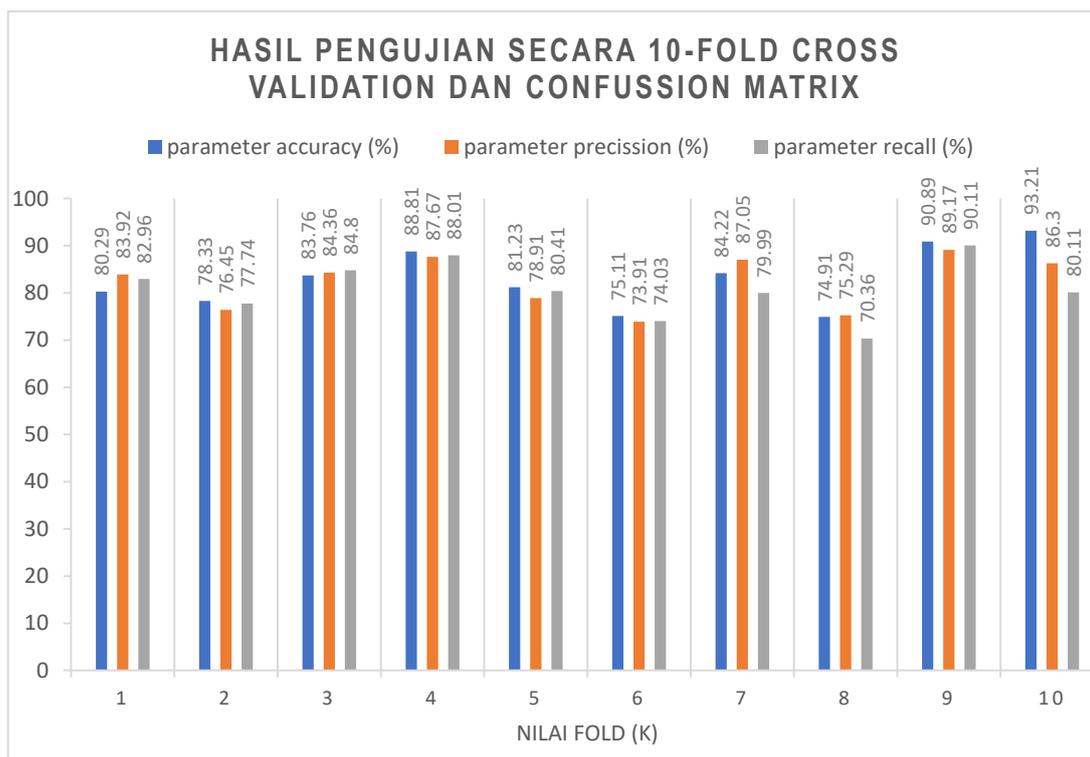
Ketika semua atribut telah dihitung probabilitasnya maka langkah selanjutnya yaitu menghitungnya dengan data uji. Rangkuman mengenai hasil klasifikasi terhadap data *testing* terdapat pada Tabel 6.

Tabel 6. Hasil klasifikasi terhadap data testing

No	A1	A2	A3	A4	...	A20	Actual class	Predicted class
1	Kepala RT	Kawin	L	71	...	3	RTSM	RTSM
2	Kepala RT	Kawin	L	80	...	2	RTSM	RTSM
3	Kepala RT	Kawin	L	52	...	4	RTSM	RTSM
4	Kepala RT	Kawin	L	50	...	3	RTSM	RTSM
5	Kepala RT	Kawin	L	39	...	3	RTSM	RTSM
6	Kepala RT	Kawin	L	50	...	3	RTSM	RTSM
7	Kepala RT	Kawin	L	30	...	4	RTM	RTM
8	Kepala RT	Kawin	L	37	...	3	RTSM	RTM
9	Kepala RT	kawin	L	69	...	2	RTM	RTSM
10	Kepala RT	Kawin	L	44	...	3	RTM	RTM
11	Kepala RT	Kawin	L	50	...	4	RTM	RTSM
12	Kepala RT	Belum Kawin	L	28	...	2	RTM	RTM
13	Kepala RT	Kawin	L	53	...	3	RTSM	RTM
14	Kepala RT	Kawin	L	48	...	3	RTM	RTM
15	Kepala RT	Kawin	L	80	...	3	RTSM	RTSM
...
110	Kepala RT	Kawin	L	73	...	3	RTM	RTSM

Untuk mengetahui akurasi yang dihasilkan oleh *naïve bayes classifier*, maka langkah selanjutnya adalah melakukan uji coba terhadap hasil yang didapatkan. Pada penelitian ini pengujian akan dilakukan dengan *confusion matrix* dan *10-fold cross validation* dimana setiap fold akan dihitung nilai *precision*, *recall*, dan *accuracy*. Rangkuman mengenai hasil pengujian *confusion matrix* terhadap *10-fold cross validation* terdapat pada Gambar 1.

Berdasarkan gambar 1 diketahui bahwa pada fold ke-1 *accuracy* yang dihasilkan adalah 80,29%; *precision* = 83,92%; dan *recall* = 82,96%. Sedangkan untuk fold ke-2 sampai dengan fold ke-10 peneliti akan menghitung nilai dari *precision*, *recall*, dan *accuracy* untuk setiap iterasinya kemudian akan dihitung nilai rata-rata yang dihasilkan dari iterasi tersebut. Dari 10-fold diketahui bahwa rata-rata akurasi tertinggi terdapat pada fold ke-10 dengan nilai *accuracy* 93,21%; *precision* 86,3%; dan *recall* 80,11%. Hal ini berarti bahwa akurasi yang dihasilkan oleh *naïve bayes classifier* dalam melakukan klasifikasi rumah tangga miskin cukup tinggi yaitu diatas 90%.



Gambar 1. Hasil pengujian 10-fold cross validation dan confusion matrix

5. KESIMPULAN

Berdasarkan hasil pengujian pada 10-fold *cross validation* yang dilakukan pada 368 dataset kemiskinan melalui survey sosial ekonomi nasional tahun 2020 di Kecamatan Gebang Kabupaten Cirebon Tahun 2020 diketahui bahwa *Naïve Bayes Classifier* mampu menunjukkan hasil yang tinggi dalam melakukan *classification* terhadap rumah tangga miskin dimasa pandemi. Hal ini terlihat pada rata-rata hasil akurasi yang dihasilkan oleh fold ke-10 di k-fold *cross validation* adalah 93,21% dengan nilai *precision* 86,3% dan *recall* 80,11%

DAFTAR PUSTAKA

- [1] W. Widayani and H. Harliana, "Perbandingan Algoritma K-Means dan SFCM Pada Pengelompokan Rumah Tangga Miskin," *J. Sains dan Inform.*, vol. 6, no. 1, pp. 1–9, 2020, doi: 10.34128/jsi.v6i1.200.
- [2] Badan Pusat Statistik Provinsi Jawa Barat, "Kemiskinan dan Ketimpangan di Jawa Barat Maret 2020," 2020.
- [3] Anonim, "Program penanggulangan kemiskinan kabinat Indonesia bersatu II," Kementerian komunikasi dan Informatika RI, Indonesia, 2016.
- [4] C. Sartika, M. Y. Balaka, and W. A. Rumbia, "Studi Faktor-Faktor Penyebab Kemiskinan Masyarakat Desa Lohia Kecamatan Lohia Kabupaten Muna," *J. Ekon.*, vol. 1, no. April, pp. 106–118, 2016.
- [5] T. Arifin and D. Ariesta, "Prediksi Penyakit Ginjal Kronis Menggunakan Algoritma Naive Bayes Classifier Berbasis Particle Swarm Optimization," *J. Tekno Insentif*, vol. 13, no. 1, pp. 26–30, 2019, doi: 10.36787/jti.v13i1.97.
- [6] H. Muhamad, C. A. Prasojo, N. A. Sugianto, L. Surtiningsih, and I. Cholissodin, "Optimasi Naïve Bayes Classifier Dengan Menggunakan Particle Swarm Optimization Pada Data Iris," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 4, no. 3, p. 180, 2017, doi: 10.25126/jtiik.201743251.

-
- [7] H. Annur, "Klasifikasi Masyarakat Miskin Menggunakan Metode Naive Bayes," *Ilk. J. Ilm.*, vol. 10, no. 2, pp. 160–165, 2018.
- [8] M. Rasyida, "Naïve Bayes Classification untuk Penentuan Status Penduduk Miskin," *J. Inform. Kaputama*, vol. 4, no. 2, 2020, doi: 10.1234/jik.v4i2.303.
- [9] A. C. Putri, F. E. Hariyanto, N. L. E. Andini, and Z. C. S. Zulkarnaen, "Klasifikasi Rumah Tangga Miskin di Provinsi Papua Tahun 2017 Menggunakan Metode Naive Bayes," *J. Sains Mat. dan Stat.*, vol. 7, no. 1, p. 89, 2021, doi: 10.24014/jsms.v7i1.11924.
- [10] N. Alfiah, "Klasifikasi Penerima Bantuan Sosial Program Keluarga Harapan Menggunakan Metode Naive Bayes," *J. Teknol. Inf.*, vol. XVI, pp. 32–40, 2021.
- [11] F. K. Pratama, D. W. Widodo, and N. Shofia, "Implementasi Metode Naive Bayes dalam Mengklasifikasi Penerima Program Keluarga Harapan (PKH) Desa Minggiran Kediri," *Semin. Nas. Inov. Teknol. UN PGRI Kediri*, pp. 23–28, 2021.
- [12] M. A. Maricar and Dian Pramana, "Perbandingan Akurasi Naive Bayes dan K-Nearest Neighbor pada Klasifikasi untuk Meramalkan Status Pekerjaan Alumni ITB STIKOM Bali," *J. Sist. dan Inform.*, vol. 14, no. 1, pp. 16–22, 2019, doi: 10.30864/jsi.v14i1.233.
- [13] M. F. Fibrianda and A. Bhawiyuga, "Analisis Perbandingan Akurasi Deteksi Serangan Pada Jaringan Komputer Dengan Metode Naive Bayes Dan Support Vector Machine (SVM)," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. II, no. 9, pp. 3112–3123, 2018.
- [14] A. F. Watratan, A. Puspita, and D. Moeis, "Implementasi Algoritma Naive Bayes Untuk Memprediksi Tingkat Penyebaran Covid-19 Di Indonesia," *J. Appl. Comput. Sci. Technol.*, vol. 1, no. 1, pp. 7–14, 2020, doi: 10.52158/jacost.v1i1.9.

Biodata Penulis

Harliana, ST., M.Cs, lahir di Cirebon pada tahun 1986. Penulis pertama memperoleh gelar S.T di Jurusan Teknik Informatika Universitas Muhammadiyah Cirebon pada Tahun 2008. Kemudian melanjutkan pendidikan S2 pada Universitas Gadjah Mada Yogyakarta dan lulus pada Tahun 2012. Konsentrasi penelitian yang penulis pertama dalam yaitu bidang *Artificial Intellience*, data mining, dan *image processing*. Saat ini penulis adalah salah satu dosen di program studi S1 Ilmu Komputer Universitas Nahdlatul Ulama Blitar.

Fatra Nonggala Putra, S.Pd., M.Kom., lahir di Kabupaten Blitar pada tanggal 01 Nopember 1990. Penulis menyelesaikan pendidikan Sarjana (S1) di Universitas Negeri Malang (UM), Jurusan Teknik Elektro, Fakultas Teknik pada Program Studi Pendidikan Teknik Informatika (PTI). Pada tahun 2016 peneliti melanjutkan pendidikan Pascasarjana (S2) di Institut Teknologi Sepuluh Nopember Surabaya (ITS). Pada Tahun 2019 hingga sekarang, penulis merupakan Dosen di program studi S1 Ilmu Komputer Universitas Nahdlatul Ulama Blitar.