

KLASIFIKASI MASYARAKAT PENERIMA BANTUAN LANGSUNG TUNAI DANA DESA MENGGUNAKAN NAÏVE BAYES DAN SMOTE

Dede Kurniadi^{*1}, Fitri Nuraeni², Marshal Firmansyah³

^{1,2,3}Institut Teknologi Garut, Kabupaten Garut
Email: ¹dede.kurniadi@itg.ac.id, ²fitri.nuraeni@itg.ac.id, ³1806130@itg.ac.id
^{*}Penulis Korespondensi

(Naskah masuk: 30 Juli 2022, diterima untuk diterbitkan: 11 April 2023)

Abstrak

Pemerintah menyelenggarakan program Bantuan Langsung Tunai Dana Desa (BLT DD), program ini memberikan (subsidi) kepada keluarga miskin yang memenuhi syarat. Program ini dapat membantu mengurangi beban pengeluaran serta meningkatkan pendapatan keluarga miskin. Masyarakat yang berhak menerima BLT DD terkadang melebihi kuota yang tersedia, kemudian proses penentuan penerima dilakukan secara musyawarah. Hasil penetapan tersebut terkadang menimbulkan kecemburuan sosial di masyarakat, sehingga diperlukan klasifikasi yang dapat membantu menentukan keluarga yang layak menerima program bantuan ini. Penelitian ini bertujuan untuk menerapkan metode Naïve Bayes untuk mengklasifikasikan data keluarga layak dan tidak layak menerima BLT DD karena masih banyak keluarga miskin berpenghasilan rendah lainnya yang belum berkesempatan untuk memperoleh program bantuan ini. Metode penelitian yang digunakan yaitu *Cross-Industry Standard Process For Data Mining* (CRISP-DM). Data yang digunakan merupakan data penerima BLT DD tahun 2021 dan 2022 di Desa Kersamenak dengan jumlah data yang digunakan sebanyak 375, meliputi *class* layak 205 *record* dan tidak layak 170 *record*. Data yang terkumpul menunjukkan adanya ketidakseimbangan kelas pada jumlah masyarakat yang layak dan tidak layak, sehingga diperlukan teknik *Synthetic Minority Over-sampling Technique* (SMOTE) untuk menangani kelas yang tidak seimbang pada data. Hasil pemodelan Naïve Bayes menggunakan teknik SMOTE menghasilkan model performansi terbaik dengan nilai akurasi 97,80% dan nilai AUC 0,99 yang termasuk dalam kategori *Excellent Classification*. Berdasarkan hasil model kinerja klasifikasi yang diperoleh, model yang dihasilkan dapat diimplementasikan ke dalam sistem aplikasi pendukung keputusan untuk membantu Desa dalam menentukan penerima BLT DD agar lebih cepat dan mudah.

Kata kunci: *Bantuan Langsung Tunai, CRISP-DM, Klasifikasi, Naïve Bayes, SMOTE*

CLASSIFICATION OF SOCIETY RECIPIENTS OF BANTUAN LANGSUNG TUNAI DANA DESA USING NAÏVE BAYES AND SMOTE

Abstract

The government organizes the Bantuan Langsung Tunai Dana Desa (BLT DD) program, which provides (subsidies) to low-income families who meet the requirements. This program can help reduce the burden of spending and increase the income of low-income families. Communities who deserve to receive BLT DD sometimes exceed the available quota, then the process of determining the recipient is carried out utilizing deliberation. The results of these determinations sometimes cause social jealousy in the community, so a classification is needed that can help determine eligible families to receive this assistance program. This study aims to apply the Naïve Bayes method to classify family data as eligible and not eligible to receive BLT DD because there are still many other low-income families who have not had the opportunity to acquire this assistance program. The research method used is *Cross-Industry Standard Process for Data Mining* (CRISP-DM). The data used is the data of the 2021 and 2022 Village Fund Direct Cash Aid recipients in Kersamenak Village, with the amount of data used as much as 375, including 205 eligible class records and 170 inappropriate records. The data collected shows an imbalanced class in the number of eligible and ineligible people, and it is necessary to use the *Synthetic Minority Over-sampling Technique* (SMOTE) technique to handle the imbalanced class in the data. The results of modeling the Naïve Bayes using SMOTE technique produce the best performance model with an accuracy value of 97.80% and an AUC value of 0.99, which is included in the *Excellent Classification* category. Based on the results of the classification performance model obtained, we can implement the resulting model into a decision support application system to assist the Village in determining the recipient of the BLT DD to make it faster and easier.

Keywords: *Bantuan Langsung Tunai, CRISP-DM, Classification, Naïve Bayes, SMOTE*

1. PENDAHULUAN

Rendahnya kualitas hidup masyarakat miskin dapat menyebabkan rendahnya tingkat pendidikan dan kesehatan, yang dapat berdampak pada produktivitas (Nul et al., 2021). Bantuan Langsung Tunai Dana Desa (BLT DD) yaitu suatu program bantuan pemerintah yang dalam bentuk uang tunai maupun bantuan lainnya baik bersyarat atau tidak bersyarat terutama diberikan kepada masyarakat miskin. Mekanisme penyaluran dan besaran dana BLT DD tergantung kepada kebijakan yang ada. Mekanisme yang digunakan dalam penyelenggara di Indonesia berupa pemberian uang tunai, pangan, jaminan kesehatan dan pendidikan (Hariandja & Budiman, 2021).

Penerima program merupakan keluarga miskin yang telah terdaftar di pemerintahan desa, yang nantinya data tersebut dikirim sampai di Kementerian Sosial. Penerima bantuan yang ada di desa terkadang melebihi kuota yang tersedia, sehingga diperlukannya pengolahan tambahan yaitu mengklasifikasi penerima program BLT DD. Pengolahan lebih lanjut dilakukan oleh Rukun Tetangga (RT), Rukun Warga (RW) dan staf bantuan sosial yang ada di desa. Prosedur penerima bantuan dimulai dengan pemberian surat keputusan dari pusat dan kemudian dikonfirmasi dengan kriteria yang telah ditentukan dalam Peraturan Menteri Keuangan Republik Indonesia nomor 190/PMK.07/2021 (pasal 33) (Keuangan, 2021), salah satunya ditinjau dari syarat BLT DD yang dituju. Proses penentuan penerima dilakukan dengan cara musyawarah, hasil penentuan tersebut terkadang menimbulkan kecemburuan sosial ditengah masyarakat sehingga perlu adanya klasifikasi yang dapat membantu untuk menentukan keluarga yang layak menerima BLT DD. Menurut Dinda & Agus yang dikutip pada jurnal penelitian (Sinaga et al., 2021) bahwa klasifikasi itu menemukan sebuah *record* data baru di salah satu dari beberapa katagori (*class*) yang telah ditentukan sebelumnya. Maka dari itu pemodelan klasifikasi ini cocok untuk menentukan penerima BLT DD.

Pada penelitian ini terdapat beberapa jurnal yang menjadi rujukan yaitu penelitian yang dilakukan (Irmayansyah & Firdaus, 2019) mengenai penentuan Bantuan Langsung Tunai kepada penduduk menggunakan algoritma *C4.5* dengan pemodelan klasifikasi memperoleh nilai akurasi 74,9%. Penelitian selanjutnya oleh (Wintana et al., 2019) mengenai klasifikasi perhitungan atribut yang menentukan penerima manfaat keluarga harapan adalah kepemilikan anak sekolah dengan nilai 0,512716784. Kemudian pada penelitian (Putri et al., 2021) mengenai klasifikasi penerima manfaat bantuan non tunai kartu keluarga sejahtera, menggunakan dua metode yaitu *Naïve Bayes* dan *K-Nearest Neighbor* bahwa penggunaan *Naïve Bayes*

memperoleh model *performance* terbaik dengan nilai akurasi sebesar 99.89%. Pada penelitian (Nuraeni et al., 2021) mengenai ketepatan masa studi mahasiswa dengan menerapkan fitur *Forward Selection* pada *Naïve Bayes* menghasilkan nilai akurasi yang cukup tinggi yaitu 92,94%. Pada penelitian (Lefya et al., 2022) menunjukkan bahwa sistem uji kelayakan penerima Bantuan Langsung Tunai (BLT) Badan Penyelenggara Jaminan Sosia (BPJS) dengan menggunakan metode klasifikasi *naïve bayes* dapat membantu untuk menyampaikan informasi berupa kelayakan seorang penerima bantuan BLT BPJS dengan hasil akurasi sebesar 83,689%. Sedangkan pada penelitian (Anwar Pauji et al., 2022) mengklasifikasi penerima manfaat BLT menggunakan metode *K-Nearest Neighbor* sehingga penerima bantuan benar-benar tepat sasaran dengan perolehan nilai akurasi sebesar 68,82 %.

Dari beberapa jurnal rujukan penelitian sebelumnya menunjukkan pemodelan menggunakan metode *Naïve Bayes* oleh (Putri et al., 2021) terbukti menghasilkan tingkat akurasi yang tinggi sebesar 99.89%. Sedangkan pada penelitian (Fitriani, 2020) metode *Naïve Bayes* menghasilkan model dengan kategori *excellent classification*, sehingga metode *Naive Bayes* dapat digunakan pada model klasifikasi dalam menentukan penerima BLT DD.

Berdasarkan data yang diperoleh terdapat *imbalanced class* antara masyarakat yang layak dan tidak layak dalam menerima BLT DD, maka diperlukan *Synthetic Minority Over Sampling Technique* (SMOTE) yang terbukti dapat menyeimbangkan data dan meningkatkan model *performance* berdasarkan penelitian yang dilakukan oleh (Siringoringo, 2018). Dari latar belakang dan kajian penelitian sebelumnya, maka penelitian ini menggunakan metode *Naïve Bayes* yang dikenal memiliki tingkat akurasi yang tinggi dengan perhitungan sederhana. Hasil dari pemodelan klasifikasi ini berupa kelayakan penerima BLT DD. Penelitian ini diharapkan dapat memudahkan bagi desa dalam menentukan penerima BLT DD sehingga orang yang menerima bantuan ini merupakan orang yang paling membutuhkan.

2. METODE PENELITIAN

Penelitian ini menggunakan metode *Cross Industry Standard Process for Data Mining* (CRISP-DM) adalah standar proses penambangan data untuk melakukan analisis dari suatu industri yang mendasari strategi pemecahan masalah perusahaan atau unit penelitian (Adiana et al., 2018).

2.1. Naïve Bayes

Naïve Bayes adalah suatu algoritma atau metode yang dapat mengklasifikasikan suatu variable tertentu dengan menggunakan metode probabilitas dan statistik. Adapun proses perhitungan pemodelan Naïve Bayes yaitu sebagai berikut (Harimurti & Riksakomara, 2017):

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

Keterangan:

- X = Sampel data dengan class yang belum diketahui
- H = Hipotesa X merupakan data class spesifik
- P(H|X) = Peluang dari hipotesa H berdasarkan kondisi X
- P(X|H) = Peluang X berdasarkan kondisi hipotesa H
- P(H) = Peluang dari hipotesa H
- P(X) = Peluang dari data sampel yang diamati

2.2. Synthetic Minority Oversampling Technique

Synthetic Minority Oversampling Technique (SMOTE) merupakan teknik yang diterapkan untuk menangani kelas yang tidak seimbangan. Teknik ini berkerja dengan cara mengambil sampel data baru. Jumlah data sampel yang diambil menyesuaikan dengan jumlah data minoritas. Teknik ini dilakukan dengan cara sampling ulang sampel class, agar perbandingan jumlah data menjadi seimbang (Siringoringo, 2018).

2.3. Uji Performa

Uji performa menggunakan Confusion Matrix untuk mengevaluasi data menggunakan Tabel matriks (Harimurti & Riksakomara, 2017) kemudian untuk mengetahui kategori klasifikasi menggunakan kurva ROC (Ardiyansyah et al., 2018). Pengujian ini menghasilkan performa model berupa nilai, berikut merupakan persamaan dalam memperoleh nilai dari hasil pengujian:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{2}$$

$$Precision = \frac{TP}{TP+FP} \tag{3}$$

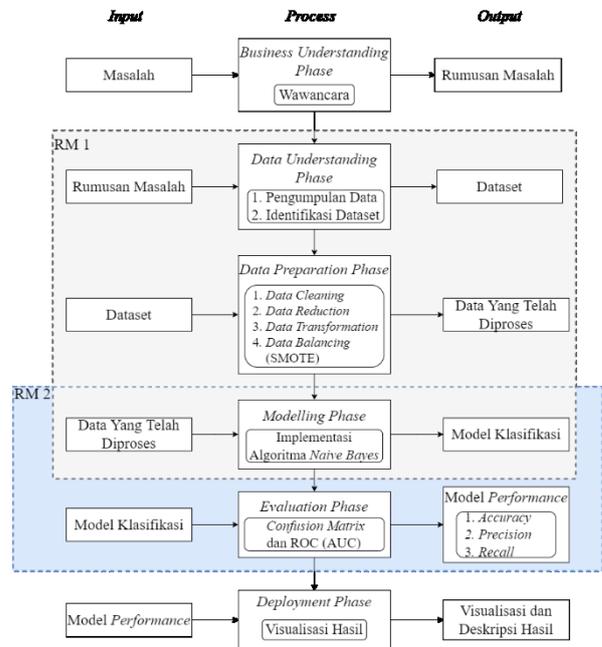
$$Recall = \frac{TP}{TP+FN} \tag{4}$$

Keterangan:

- P : Kondisi Positif
- N : Kondisi Negatif
- TP : Benar Positif
- TN : Benar Negatif
- FP : Salah Positif
- FN : Salah Negatif

2.4. Kerangka Penelitian

Kerangka penelitian adalah diagram yang menggambarkan alur berjalannya suatu penelitian. Kerangka penelitian ini disajikan dengan menggunakan pendekatan metode CRISP-DM sebagai berikut:



Gambar 1. Kerangka Penelitian

Keterangan:

- RM 1 : “Bagaimana mengklasifikasi masyarakat penerima BLT DD menggunakan metode Naïve Bayes untuk mengetahui nilai probabilitas setiap atribut?” Berdasarkan pertanyaan Rumusan Masalah 1 pada penelitian ini, akan terjawab pada tahapan Data Understanding Phase & Data Preparation Phase;
- RM 2 : “Bagaimana menerapkan metode Naïve Bayes dalam penerima BLT DD dengan tingkat akurasi yang baik?” Berdasarkan pertanyaan Rumusan Masalah 2 pada penelitian ini, akan terjawab pada tahapan Modelling Phase & Evaluation Phase.

Berdasarkan Gambar 1 terdapat 3 tahapan yaitu *input*, *process* dan *output*. *Input* merupakan sumber yang diperlukan untuk menghasilkan *output*. *Process* merupakan kegiatan mengolah *input* agar mendapatkan *output*. Sedangkan *output* merupakan hasil yang diperoleh dari *input* yang telah di *process*. Pada tahapan *process* terdiri dari 6 tahapan metode CRISP-DM yaitu *Business Understanding Phase*, *Data Understanding Phase*, *Data Preparation Phase*, *Modelling Phase*, *Evaluation Phase* dan *Deployment Phase*.

2.5. Sumber Data

Sumber data yang digunakan untuk kebutuhan analisis data dan pelatihan model dalam penelitian ini bersumber dari Desa Kersamenak. Data yang digunakan merupakan data Bantuan Langsung Tunai Dana Desa tahun 2021 dan 2022. Hasil perolehan data Bantuan Langsung Tunai Dana Desa mengenai status kelayakan yang layak menerima bantuan sebanyak 205 orang dan tidak layak sebanyak 170 orang dengan jumlah data yang terkumpul sebanyak 375 data. Atribut yang digunakan sebanyak 17 jenis yaitu NIK, Nama, Usia, Status Pernikahan, Status Keluarga, Pendidikan Terakhir, Pekerjaan, RT, RW, Alamat, Kemiskinan Ekstrem, Kehilangan Mata Pencarian, Anggota Keluarga Yang Rentan Sakit, Bantuan Sosial Yang Berhenti, Terdampak Pandemi, Rumah Tangga Tunggal Lanjut Usia, dan Status Kelayakan.

3. HASIL DAN PEMBAHASAN

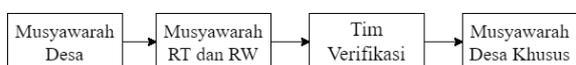
Hasil dan pembahasan dari penelitian ini merupakan penerapan dari metode *Naïve Bayes* menggunakan teknik SMOTE. Pemodelan ini dibantu dengan *tools* tambahan yaitu *Rapidminer*. Untuk hasil penelitian dan pembahasan hasil yang diperoleh disajikan pada sub bahasan berikut:

3.1. Hasil Penelitian

Penelitian ini menggunakan metode pendekatan CRISP-DM. Adapun hasil dari penelitian ini akan dijelaskan berdasarkan tahapan CRISP-DM. Untuk lebih jelasnya dapat dilihat pada sub poin berikut:

3.1.1. Business Understanding Phase

Berdasarkan objek penelitian yang telah ditentukan, maka tahapan pertama dari penelitian ini yaitu dengan melakukan aktifitas wawancara. Aktifitas wawancara adalah diskusi secara langsung maupun tidak langsung untuk menemukan sebuah permasalahan serta mendapatkan solusi dari permasalahan tersebut. Dalam melakukan wawancara terdapat permasalahan bahwa ada batasan kuota bagi penerima Bantuan Langsung Tunai Dana Desa (BLT DD) sehingga masyarakat yang membutuhkan bantuan tidak semuanya menerima program bantuan ini. Maka dari permasalahan tersebut perlu adanya klasifikasi yang dapat membantu menentukan keluarga yang layak untuk mendapatkan BLT DD. Berdasarkan wawancara yang telah dilakukan, terdapat alur dalam penerimaan BLT DD Desa yang tertera pada Gambar 2 berikut:



Gambar 2. Alur BLT DD

Berdasarkan Gambar 2 terdapat alur dalam penerimaan BLT DD yaitu sebagai berikut:

1. Musyawarah Desa (Musdes)

Membahas tentang pembentukan tim verifikasi untuk menentukan calon Keluarga Penerima Manfaat (KPM) dan menentukan jumlah KPM disesuaikan dengan pagu BLT DD yang disediakan oleh Desa. Besaran penyaluran BLT DD ditetapkan sebesar Rp 300.000 untuk bulan pertama sampai bulan kedua belas untuk setiap KPM.
2. Musyawarah RT dan RW

Penugasan kepada setiap RT dan RW untuk mengajukan calon Penerima BLT DD sesuai dengan kriteria yang telah ditentukan dalam Peraturan Menteri Keuangan Republik Indonesia nomor 190/PMK.07/2021 (pasal 33) (Keuangan, 2021).
3. Tim Verifikasi

Setelah pengajuan calon Penerima BLT DD oleh RT dan RW maka data tersebut akan di verifikasi mengenai kelayakan apakah keluarga tersebut layak menerima atau tidak dalam BLT DD termasuk menyesuaikan jumlah kuota yang telah ditentukan pada Musyawarah Desa.
4. Musyawarah Desa Khusus (Musdesus)

Penetapan Keluarga Penerima Manfaat penerima BLT DD dari hasil tim verifikasi yang telah mendata calon penerima dengan melibatkan lembaga kemasyarakatan Desa.

3.1.2. Data Understanding Phase

Tahap ini adalah pengumpulan data yang kemudian akan dilanjutkan melalui proses untuk mendapatkan pemahaman yang mendalam tentang data dan mengidentifikasinya. Adapun Langkah-langkah pada tahapan ini sebagai berikut:

1. Tahap pengumpulan data yaitu proses mencari data yang akan digunakan untuk penelitian ini. Data yang diperoleh merupakan data penerima BLT DD 2021 dan 2022 yang ada di Desa Kersamenak. Berikut data yang sudah terkumpul terdapat pada Tabel 1:

Tabel 1. Data Mentah

No	NIK	Nama	...	Status Kelayakan
1	32050525 1207****	Oya	...	Layak
2	32050518 1045****	Aja	...	Tidak Layak
3	32050526 1207****	Nunuh	...	Layak
...
375	32050570 0194****	Indri Purnama Sari	...	Tidak Layak

Data pada Tabel 1 merupakan data yang masih mentah yang berjumlah sebanyak 375 *record* dan terdiri dari 17 atribut yaitu NIK, Nama, Usia, Status Pernikahan, Status

Keluarga, Pendidikan Terakhir, Pekerjaan, RT, RW, Alamat, Kemiskinan Ekstrem, Kehilangan Mata Pencarian, Anggota Keluarga Yang Rentan Sakit, Bantuan Sosial Yang Berhenti, Terdampak Pandemi, Rumah Tangga Tunggal Lanjut Usia, dan Status Kelayakan.

2. Tahap mengidentifikasi jenis dataset dan menentukan atribut yang akan digunakan. Tipe atribut dataset yang diperoleh pada penelitian ini terdapat pada Tabel 2 sebagai berikut:

Tabel 2. Tipe Atribut Dataset

No	Atribut	Tipe Data	Parameter
1	NIK	Integer	-
2	Nama	Polynomial	-
4	Usia	Integer	-
5	Status Pernikahan	Polynomial	Belum Kawin, Kawin, Cerai Hidup, Cerai Mati
...
17	Status Kelayakan	Binominal	Layak dan Tidak Layak

Berdasarkan pada Tabel 2 terdapat 17 atribut yang terdiri dari 4 tipe data *integer* pada atribut NIK, Usia, RT dan RW. Terdapat 4 data *polynomial* pada atribut Nama, Status Pernikahan, Pendidikan Akhir dan Alamat. Terdapat 8 tipe data *binominal* pada atribut Status Keluarga, Kemiskinan Ekstrem, Kehilangan Mata Pencarian, Anggota Keluarga yang Rentan Sakit, Bantuan Sosial yang Berhenti, Terdampak Pandemi, Rumah Tangga Tunggal Lanjut Usia dan Status Kelayakan.

3.1.3. Data Preparation Phase

Data mentah yang diperoleh pada tahapan sebelumnya perlu dilakukan *Data Preparation Phase*. Proses yang harus dilakukan dalam pengolahan data pada penelitian ini adalah sebagai berikut:

1. *Data cleaning* yaitu proses pembersihan data jika terdapat data yang tidak valid, kosong dan kurang relevan pada atribut tertent. Berdasarkan data yang terkumpul, terdapat 2 *record* yang kosong pada parameter pendidikan terakhir maka perlu dilakukan proses *cleaning*. Jumlah data awal sebanyak 375 *record* dan setelah dilakukan proses *cleaning* menjadi 373 *record*.
2. *Data Reduction* yaitu menentukan atau memilih atribut mana yang akan digunakan. Dari data yang terkumpul terdapat atribut yang tidak digunakan, untuk proses *data reduction* terdapat pada Gambar 3 berikut:



Gambar 3. Hasil *Data Reduction*

Berdasarkan Gambar 3 proses *data reduction* terdapat beberapa atribut yang tidak digunakan seperti NIK yang bersifat privasi oleh karena itu NIK tidak digunakan sebagai Id, maka atribut Nama yang menjadi Id nya. Sedangkan atribut RT dan RW tidak digunakan karena atribut tersebut bagian dari alamat.

3. Tahap selanjutnya yaitu melakukan *data transformation*. Proses *transformation* yang dilakukan yaitu dengan merubah parameter pada atribut. Berikut proses *data transformation* yang terdapat pada Tabel 3:

Tabel 3. Hasil *Data Transformation* Parameter

No	Atribut	Parameter (Sebelum di Transformation)	Parameter (Sesudah di Transformation)
1	Usia	26-35 tahun	Dewasa Awal
		36-45 tahun	Dewasa Akhir
		46-55 tahun	Lansia Awal
		56-65 tahun	Lansia Akhir
		66-100 tahun	Manula
2	Pendidikan Terakhir	Tidak/Belum Sekolah	Tidak Sekolah
		Tidak Tamat SD/Sederajat	Tidak Tamat SD
		Tamat SD/Sederajat	Tamat SD/
		SLTP/Sederajat	SMP
3	Jarak Alamat Ke Kota / KM	SLTA/Sederajat	SMA
		Kp. Cirengit	5,4
		Kp. Babakan	5,8
		Kp. Salamgede	6,2
		Kp. Ancol	4,9
		Kp. Nangewer	4,7
		Kp. Cicurug	5
		Kp. Kiarapayung	6
		Kp.Pameungpeuk	5,5
		Kp. Ganea	5,2
Kp. Cicadas	4,8		

Berdasarkan Tabel 3 proses *data transformation* yang dilakukan dengan merubah parameter pada atribut usia, pendidikan terakhir dan jarak alamat ke kota /km untuk memudahkan pada proses *modelling phase*.

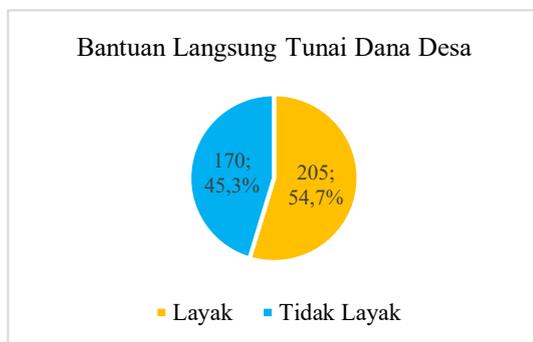
Proses *data transformation* ini juga dilakukan pada atribut yang tertera pada Tabel 4 berikut:

Tabel 4. Hasil *Data Transformation* Atribut

No	Atribut (Sebelum di Transformation)	Atribut (Sesudah di Transformation)
1	Kemiskinan Ekstrem	K1
2	Kehilangan Mata Pencaharian	K2
3	Anggota Keluarga Yang Rentan Sakit	K3
4	Bantuan Sosial Yang Berhenti	K4
5	Terdampak Pandemi	K5
6	Rumah Tangga Tunggal Lanjut Usia	K6

Berdasarkan Tabel 4 proses *data transformation* dilakukan pada setiap atribut agar memudahkan pada *modelling phase*.

4. *Data Balancing* yaitu proses menyeimbangkan data menggunakan teknik SMOTE. Proses ini dilakukan karena terdapat *imbalanced class* pada masyarakat yang layak dan tidak layak. Sehingga diperlukan proses penyeimbangan data. Adapun persentase *class* yang layak dan tidak layak dapat dilihat pada Gambar 4 berikut:



Gambar 4. Diagram *Imbalanced Class*

Berdasarkan Gambar 4 terjadi *imbalanced class* pada data tersebut memiliki perbandingan yaitu 45,3% yang layak dan 54,7% tidak layak. Maka dibutuhkan proses untuk menyeimbangkan data *class* tersebut dengan menggunakan teknik SMOTE dengan cara sampling ulang data (Siringoringo, 2018).

Proses sampling ulang dilakukan terhadap jumlah data masyarakat yang tidak layak dengan menyesuaikan jumlah data masyarakat yang layak. Proses pengambilan data sampel dilakukan secara *random*, sehingga dibutuhkan *tool* yang dapat membantu dalam proses penyeimbangan data dengan menerapkan operator SMOTE Upsampling pada *tool Rapidminer*.

3.1.4. Modelling Phase

Dalam tahap ini akan dilakukan dengan menerapkan teknik pemodelan data mining, alat bantu data mining, algoritma atau metode yang digunakan dan menyesuaikan aturan model untuk mendapatkan nilai yang optimal.

Langkah pertama untuk mencari nilai probabilitas suatu *class* yaitu atribut Status Kelayakan. Adapun proses untuk menghitung probabilitas menggunakan persamaan (1) sebagai berikut:

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

$$P(\text{Layak}) = \frac{205}{373} = 0,550$$

$$P(\text{Tidak Layak}) = \frac{168}{373} = 0,450$$

Selanjutnya mencari nilai probabilitas pada setiap atribut. Adapun proses untuk menghitung probabilitas atribut Status Pernikahan menggunakan persamaan (1) yang terdapat pada Tabel 5 berikut:

Tabel 5. Nilai Probabilitas Atribut Status Pernikahan

Status Pernikahan	Class	
	Layak	Tidak Layak
Belum Kawin	0,01	0,02
Kawin	0,43	0,82
Cerai Hidup	0,05	0,05
Cerai Mati	0,51	0,11

Berdasarkan Tabel 5 perolehan nilai probabilitas tertinggi pada masyarakat layak yaitu 0,51 terdapat pada parameter ‘cerai mati’ dan nilai probabilitas terendah pada masyarakat layak yaitu bernilai 0,01 pada parameter ‘belum kawin’ dari jumlah masyarakat sebanyak 205 orang.

Sedangkan perolehan nilai probabilitas tertinggi pada masyarakat tidak layak yaitu 0,83 terdapat pada parameter ‘kawin’ dan nilai probabilitas terendah pada masyarakat tidak layak yaitu bernilai 0,02 pada parameter ‘belum kawin’ dari jumlah masyarakat sebanyak 168 orang.

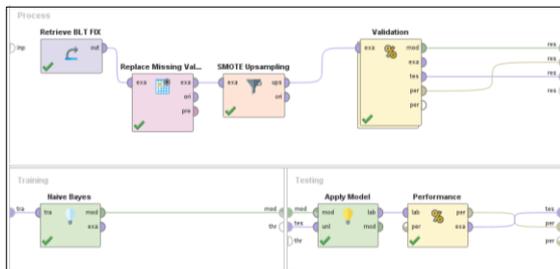
Selanjutnya yaitu melakukan proses klasifikasi dengan menghitung semua nilai probabilitas atribut. Adapun proses untuk menghitung nilai label yaitu dengan mengalikan nilai seluruh atribut label. Setelah itu cari nilai tertinggi pada label, maka itu merupakan hasil dari klasifikasi tersebut. Untuk proses perhitungan klasifikasi terdapat pada Tabel 6 berikut:

Tabel 6. Proses Perhitungan Klasifikasi

No	Nama	Usia	Status Pernikahan	...	Status Kelayakan
1	Oya_2021	Manula	Cerai Mati	...	Layak
	Layak	0,44	0,51	...	0,00006
	Tidak Layak	0,16	0,11	...	0,00000

Berdasarkan Tabel 6 proses perhitungan klasifikasi dilakukan dengan cara mengalikan seluruh nilai parameter atribut pada setiap *class* yaitu Layak dan Tidak Layak. Setelah menemukan hasilnya, lakukan perbandingan diantara kedua *class* tersebut. Berdasarkan hasil yang diperoleh, nilai tertinggi didapat dari *Class* Layak. Maka penerapan model klasifikasi dalam menentukan penerima Bantuan Langsung Tunai Dana Desa dinyatakan benar atau sesuai berdasarkan data yang dimiliki. Proses klasifikasi ini dilakukan pada seluruh data, yaitu sebanyak 373 *record*.

Pada penelitian ini menggunakan *tool* tambahan yaitu *Rapidminer* untuk membantu proses dalam pemodelan. Model klasifikasi menggunakan metode *Naive Bayes*, teknik *SMOTE* dan *k-fold cross-validation*. Berikut adalah proses pemodelan dengan *tool Rapidminer*:



Gambar 5. Pemodelan Menggunakan Tool Rapidminer

Berdasarkan Gambar 5 menunjukkan proses pemodelan menggunakan *tool Rapidminer* untuk operator yang digunakan adalah sebagai berikut: 1) *Import* dataset yang akan digunakan lalu hubungkan dengan operator *Replace Missing Values*, berfungsi untuk menormalisasikan data yang *missing value*; 2) Selanjutnya hubungkan ke operator *SMOTE Upsampling*, berfungsi untuk menyeimbangkan *Imbalanced Data*; 3) Kemudian hubungkan ke operator *Cross Validation*, berfungsi untuk mengevaluasi kinerja model atau algoritma dimana data dipisahkan menjadi dua bagian yaitu data pelatihan sementara dan data uji kinerja; 5) Lakukan double klik pada operator *Cross Validation*, pilih operator *Naive Bayes*, operator ini digunakan sebagai algoritma untuk proses pemodelan; 6) Lalu hubungkan ke operator *Apply Model*, berfungsi untuk proses pemodelan; 7) Dan terakhir hubungkan ke operator *Performance*, berfungsi untuk mengetahui performa yang dihasilkan dari proses pemodelan.

Hasil pemodelan ini diperoleh sebuah tabel distribusi yang menyatakan nilai probabilitas setiap label atribut. Pada Tabel 7 merupakan hasil nilai probabilitas setiap atribut dari penerapan metode *Naive Bayes* dan *SMOTE* dengan *tool Rapidminer*.

Tabel 7. Distribusi Nilai Probabilitas Setiap Atribut

Attribute	Parameter	Naive Bayes dan SMOTE	
		Layak	Tidak Layak
Usia	Value=Manula	0,44	0,17
Usia	Value=Dewasa Awal	0,07	0,11
Usia	Value=Lansia Akhir	0,26	0,16
Usia	Value=Dewasa Akhir	0,11	0,32
Usia	Value=Lansia Awal	0,12	0,24
Usia	Value=unknown	0,00	0,00
Status Pernikahan	Value=Cerai Mati	0,51	0,11
Status Pernikahan	Value=Kawin	0,43	0,82
Status Pernikahan	Value=Belum Kawin	0,01	0,02
Status Pernikahan	Value=Cerai Hidup	0,05	0,05
Status Pernikahan	Value=unknown	0,00	0,00
Status Pernikahan	Value=Kepala Keluarga	0,85	0,87
Status Keluarga	Value=Istri	0,15	0,13
Status Keluarga	Value=unknown	0,00	0,00
Pendidikan Terakhir	Value=Tamat SD	0,78	0,60
Pendidikan Terakhir	Value=SMP	0,11	0,16
Pendidikan Terakhir	Value=SMA	0,07	0,24
Pendidikan Terakhir	Value=Tidak Tamat SD	0,04	0,00
Pendidikan Terakhir	Value=Tidak Sekolah	0,00	0,00
Pendidikan Terakhir	Value=Unknown	0,00	0,00
Pekerjaan	Value=Buruh Harian	0,22	0,21
Pekerjaan	Value=Petani	0,01	0,07
Pekerjaan	Value=Wiraswasta	0,10	0,52
Pekerjaan	Value=Mengurus Rumah Tangga	0,26	0,14
Pekerjaan	Value=Tidak Bekerja	0,40	0,05
Pekerjaan	Value=Buruh Harian Lepas	0,00	0,00
Pekerjaan	Value=Unknown	0,00	0,00
Jarak Alamat Ke Kota /KM	Value=5.4	0,17	0,13
Jarak Alamat Ke Kota /KM	Value=5.8	0,17	0,09
Jarak Alamat Ke Kota /KM	Value=6.2	0,15	0,10
Jarak Alamat Ke Kota /KM	Value=4.9	0,11	0,10
Jarak Alamat Ke Kota /KM	Value=4.7	0,14	0,14

Attribute	Parameter	Naive Bayes dan SMOTE	
		Layak	Tidak Layak
Jarak			
Alamat Ke Kota /KM	Value=5	0,06	0,08
Jarak			
Alamat Ke Kota /KM	Value=6	0,02	0,13
Jarak			
Alamat Ke Kota /KM	Value=5.5	0,08	0,09
Jarak			
Alamat Ke Kota /KM	Value=5.2	0,05	0,06
Jarak			
Alamat Ke Kota /KM	Value=4.8	0,05	0,08
Jarak			
Alamat Ke Kota /KM	Value=Unknown	0,00	0,00
K1	Value=Ya	0,42	0,00
K1	Value=Tidak	0,58	1,00
K1	Value=Unknown	0,00	0,00
K2	Value=Ya	0,50	0,00
K2	Value=Tidak	0,50	1,00
K2	Value=Unknown	0,00	0,00
K3	Value=Ya	0,39	0,00
K3	Value=Tidak	0,61	1,00
K3	Value=Unknown	0,00	0,00
K4	Value=Ya	0,19	0,00
K4	Value=Tidak	0,81	1,00
K4	Value=Unknown	0,00	0,00
K5	Value=Ya	0,69	1,00
K5	Value=Tidak	0,31	0,00
K5	Value=Unknown	0,00	0,00
K6	Value=Ya	0,43	1,00
K6	Value=Tidak	0,57	0,00
K6	Value=Unknown	0,00	0,00

Berdasarkan Tabel 7 perolehan nilai probabilitas tertinggi pada masyarakat layak yaitu 0,85 terdapat pada parameter ‘Kepala Keluarga’ dan nilai probabilitas terendah pada masyarakat layak yaitu bernilai 0,01 pada parameter ‘Petani’ dari jumlah masyarakat sebanyak 205 orang.

Sedangkan perolehan nilai probabilitas tertinggi pada masyarakat tidak layak yaitu 1,00 terdapat pada atribut K5 dan K6 dengan parameter ‘Ya’ dan nilai probabilitas terendah pada masyarakat tidak layak yaitu bernilai 0,00 terdapat pada atribut K1, K2, K3 dan K4 dengan parameter ‘Tidak’ dari jumlah masyarakat sebanyak 168 orang.

3.1.5. Evaluation Phase

Tahapan ini merupakan evaluasi model dari hasil klasifikasi yang diperoleh pada tahapan sebelumnya. Proses yang dilakukan adalah uji performa menggunakan *Confusion Matrix* kemudian untuk mengetahui kategori klasifikasi menggunakan kurva ROC.

Untuk mengetahui model manakah yang menghasilkan *performance* terbaik, adapun hasil perbandingan dalam penerapan metode *Naive Bayes* terdapat pada Tabel 8 berikut:

Tabel 8. Perbandingan Model

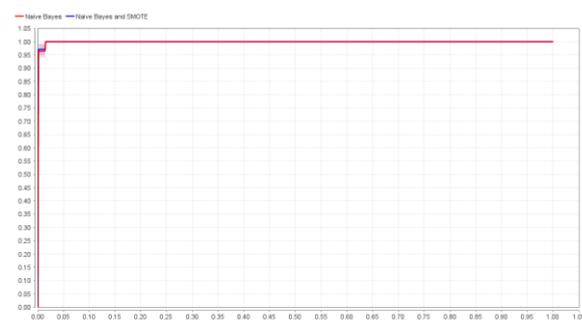
Jenis Model	Accuracy	Precision	Recall	AUC
1	97,07%	93,55%	100%	0,990
2	97,80%	96,67%	99,02%	0,999

Keterangan:

Model 1 : *Naive Bayes*

Model 2 : *Naive Bayes* dan SMOTE

Berdasarkan Tabel 8 dari beberapa model yang digunakan pada penelitian ini perolehan nilai *Recall* tertinggi didapat dari model 1 yaitu penerapan metode *Naive Bayes*. Sedangkan nilai *Accuracy*, *Precision*, dan AUC tertinggi diperoleh pada model no 2, yaitu penerapan metode *Naive Bayes* dan SMOTE. Maka dapat disimpulkan bahwa model terbaik dalam penerapan metode *Naive Bayes* yaitu terdapat pada model 2. Hasil perbandingan algoritma *Naive Bayes* disajikan dalam kurva ROC *Compare* pada Gambar 6 berikut:



Gambar 6. Kurva ROC Perbandingan Metode *Naive Bayes*

Berdasarkan Gambar 6 pada kurva ROC terdapat 2 sumbu, yaitu Sumbu X dan sumbu Y. Sumbu X merupakan *False Positive Rate* sebagai *Specificity* sedangkan sumbu Y merupakan *True Positive Rate* sebagai *Sensitivity*. Kurva perbandingan penerapan metode *Naive Bayes* pada gambar terdiri dari 2 model. Pertama garis kurva yang berwarna merah merupakan hasil dari penerapan metode *Naive Bayes* dengan perolehan nilai AUC sebesar 0,99. Sedangkan garis kurva yang berwarna biru merupakan hasil dari penerapan metode *Naive Bayes* dan teknik SMOTE dengan perolehan nilai AUC sebesar 0,99. Berdasarkan kurva ROC yang diperoleh bahwa model terbaik dalam penerapan *Naive Bayes* yaitu dengan menambahkan teknik SMOTE.

Tahapan evaluasi penelitian ini dilakukan dengan 2 metode pengujian yaitu *Confusion Matrix* meliputi *accuracy*, *precision*, *recall* dan Kurva ROC (AUC). Penggunaan *Confusion Matrix* ini bertujuan untuk menghitung performa model klasifikasi menggunakan metode *Naïve Bayes* dan teknik SMOTE. Adapun hasil yang diperoleh terdapat pada Tabel 8 berikut:

Tabel 9. Performa Model

Predicted Condition	True Condition	
	Positive	Negative
Positive	203	7
Negative	2	198

Proses untuk mengetahui nilai *Accuracy* dilakukan dengan menggunakan persamaan (2). Adapun proses perhitungannya adalah sebagai berikut:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} = \frac{203+198}{203+198+7+2} = \frac{401}{410} = 0,9780 / 97,80\%$$

Sedangkan proses untuk menghitung nilai *Precision* dilakukan dengan menggunakan persamaan (3). Adapun proses perhitungannya adalah sebagai berikut:

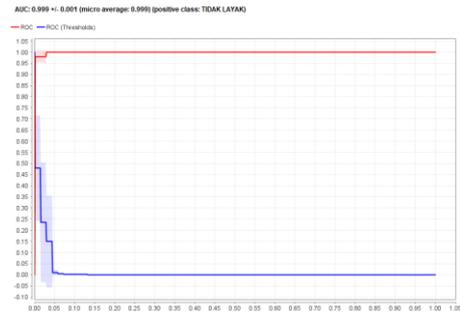
$$Precision = \frac{TP}{TP+FP} = \frac{203}{203+7} = \frac{203}{210} = 0,9667 / 96,67\%$$

Kemudian proses untuk menghitung nilai *Recall* dilakukan dengan menggunakan persamaan (4). Adapun proses perhitungannya adalah sebagai berikut:

$$Recall = \frac{TP}{TP+FN} = \frac{203}{203+2} = \frac{203}{205} = 0,9902 / 99,02\%$$

Setelah diketahui performa model dengan perolehan nilai *Accuracy* 97.80%, *Precision* 96,67% dan *Recall* 99,02%. Maka klasifikasi penerima BLT DD dengan menggunakan metode *Naïve Bayes* dan *Synthetic Minority Over Sampling Technique* (SMOTE) dinyatakan baik dan cukup akurat dikarenakan memiliki tingkat akurasi yang tinggi.

Proses evaluasi selanjutnya yaitu kurva ROC yang bertujuan untuk mengetahui kategori dan nilai dari hasil penelitian ini terdapat pada Gambar 7 sebagai berikut:

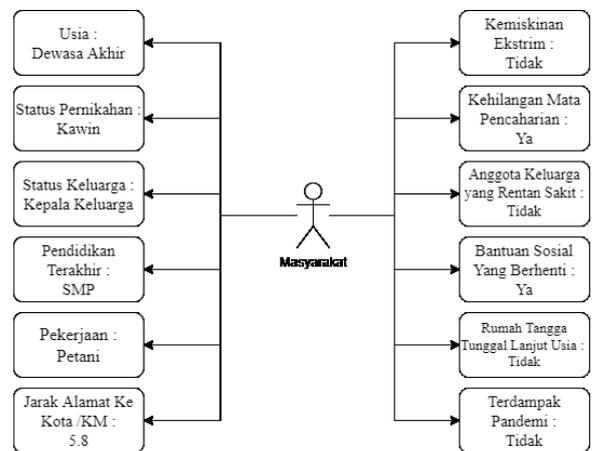


Gambar 7. Hasil Kurva ROC

Berdasarkan Gambar 7 pada kurva ROC terdapat 2 sumbu, yaitu Sumbu X dan sumbu Y. Sumbu X merupakan *False Positive Rate* sebagai *Spesitivity* sedangkan sumbu Y merupakan *True Positive Rate* sebagai *Sensitivity*. Selain itu, terdapat 2 garis pada kurva yaitu garis merah dan garis biru. Garis merah pada gambar merupakan representasi data *training* sedangkan garis biru merupakan representasi data *testing*. Jika semakin tinggi *True Positive Rate* dan semakin kecil *False Positive Rate* maka *Thresholdnya* semakin bagus. Untuk membandingkan nilai performa kurva merah dan kurva biru dalam bentuk angka dilakukan dengan membandingkan luas daerah di bawah kurva atau *Area Under Curve* (AUC). Dapat dilihat bahwa luas daerah kurva berwarna merah pada gambar tersebut lebih besar dibanding luas kurva berwarna biru dan menghasilkan nilai AUC sebesar 0,99 maka termasuk sebagai kategori *Excellent Classification* berdasarkan tabel nilai AUC (Ardiyansyah et al., 2018; Kurniadi et al., 2021).

3.1.6. Deployment Phase

Tahapan ini menjadi gambaran dari hasil model klasifikasi yang nantinya akan digunakan oleh pihak Desa dalam menentukan penerima Bantuan Langsung Tunai Dana Desa (BLT DD). Untuk ilustrasi penerapan model klasifikasi penerima BLT DD pada Gambar 8 sebagai berikut:



Gambar 8. Ilustrasi Penerapan Model Klasifikasi

Berdasarkan Gambar 8 ilustrasi penerapan model klasifikasi dalam penerimaan program BLT DD di Desa Kersamenak. Data yang tertera pada gambar merupakan parameter atribut yang dimiliki masyarakat tersebut. Untuk mengetahui masyarakat tersebut apakah layak atau tidak layak dalam menerima BLT DD maka diperlukan proses klasifikasi. Proses klasifikasi yang dilakukan yaitu dengan cara menghitung nilai probabilitas setiap atribut yang sudah diperoleh pada tabel distribusi dan membagi 2 *class* yaitu Layak dan Tidak Layak. Untuk proses perhitungan klasifikasi dalam menentukan penerima BLT DD terdapat pada Tabel 10 sebagai berikut:

Tabel 10. Proses Klasifikasi Menentukan Penerima BLT DD

Data Masyarakat	Dewa sa Akhir	Kawin	Kepala Keluarga	...	Layak
Layak	0,11	0,43	0,85	...	0,0000024
Tidak Layak	0,32	0,82	0,87	...	0,0000000

Berdasarkan Tabel 10 proses perhitungan klasifikasi dilakukan dengan cara mengalikan seluruh nilai parameter atribut pada setiap *class*. Setelah menemukan hasilnya, lakukan perbandingan diantara kedua *class* tersebut. Berdasarkan hasil yang diperoleh, nilai tertinggi didapat dari *class* Layak. Maka ilustrasi penerapan model klasifikasi untuk yang terdapat pada *Modelling Phase* menentukan penerima BLT DD dinyatakan benar atau sesuai berdasarkan data yang dimiliki.

3.2. Pembahasan Hasil

Hasil penelitian yang telah dilakukan telah berhasil menjawab 2 rumusan permasalahan yaitu: 1) klasifikasi masyarakat penerima BLT DD menggunakan metode *Naïve Bayes* untuk mengetahui nilai probabilitas setiap atribut, proses tersebut dilakukan pada tahapan *modelling*. Untuk hasil perhitungan nilai probabilitas setiap atribut terdapat pada Tabel 6. 2) Penerapan metode *Naïve Bayes* dan teknik SMOTE menghasilkan nilai *Accuracy* sebesar 97,80%, nilai *Precision* sebesar 96,67% dan *Recall* sebesar 99,02% dengan nilai AUC sebesar 0,99 termasuk pada kategori *Excellent Classification*.

Keselarasan hasil penelitian ini bahwa penerapan *Naïve Bayes* menghasilkan nilai akurasi yang tinggi hal ini selaras dengan penelitian (Putri et al., 2021) bahwa metode *Naïve Bayes* memperoleh model *performance* terbaik. Selain itu, *Naïve Bayes* juga telah berhasil menghasilkan nilai AUC tergolong *excellent classification* selaras dengan penelitian (Fitriani, 2020) bahwa *Naïve Bayes* menghasilkan nilai AUC yang tergolong sangat baik. Hasil penelitian ini juga selaras dengan penelitian (Siringoringo, 2018) dimana penggunaan teknik SMOTE terbukti dapat menyeimbangkan *class* dan mempengaruhi hasil tingkat akurasi. Kemudian keselarasan pada penelitian yang dilakukan oleh

(Kurniadi et al., n.d.) pada objek penelitian yang sama yaitu membahas mengenai penerima bantuan sosial.

Kontribusi penelitian ini bahwa pemodelan metode *Naïve Bayes* dan teknik SMOTE menghasilkan model *performance* terbaik, sehingga kedepannya dapat diimplementasikan pada sebuah sistem pendukung keputusan untuk membantu proses penentuan BLT DD agar lebih cepat dan mudah.

4. KESIMPULAN

Berdasarkan hasil penelitian yang diperoleh, maka dapat ditarik kesimpulan yaitu: 1) Metode *Naïve Bayes* telah berhasil menghasilkan nilai probabilitas setiap atribut. Nilai probabilitas ini digunakan untuk menghitung besar peluang dalam suatu kejadian. Nilai probabilitas tertinggi untuk label layak terdapat pada atribut K5 dan K6 dengan parameter 'Ya' sebesar 1,00 sedangkan nilai probabilitas terendah terdapat pada atribut K1, K2, K3 dan K4 dengan parameter 'Tidak' sebesar 0,00. 2) Hasil Evaluasi model klasifikasi dengan penerapan *Naïve Bayes* tanpa teknik SMOTE memperoleh nilai *Accuracy* sebesar 97,07%. Sedangkan penerapan *Naïve Bayes* dan teknik SMOTE dapat meningkatkan nilai akurasi yang diperoleh menjadi 97,80%. Berdasarkan nilai akurasi yang diperoleh dengan menggunakan teknik SMOTE dapat mempengaruhi terhadap tingkat akurasi. 3) Hasil perolehan nilai AUC pada visualisasi kurva ROC berdasarkan penelitian yang dilakukan menghasilkan kinerja model sebesar 0,99 dan termasuk pada kategori *Excellent Classification*.

Saran untuk penelitian selanjutnya yaitu: 1) Mengimplementasikan hasil pemodelan kedalam sistem untuk membantu proses penentuan penerima BLT DD agar lebih cepat dan mudah; 2) Dalam proses klasifikasi sebaiknya menggunakan dataset yang lebih banyak lagi atau menambahkan data dari Desa lainnya; 3) Dapat menggunakan algoritma atau metode lain untuk mengklasifikasikan penerima BLT DD seperti *K-Nearest Neighbors* (KNN) atau *Decision Tree*.

DAFTAR PUSTAKA

- ADIANA, B. E., SOESANTI, I., & PERMANASARI, A. E. 2018. Analisis Segmentasi Pelanggan Menggunakan Kombinasi RFM Model Dan Teknik Clustering. *Jurnal Terapan Teknologi Informasi*, 2(1), 23–32. <https://doi.org/10.21460/jutei.2018.21.76>
- ANWAR PAUJI, AISYAH, S., SURIP, A., SAPUTRA, R., & ALI, I. 2022. Implementasi Algoritma K-Nearest Neighbor Dalam Menentukan Penerima Bantuan Langsung Tunai. *KOPERTIP: Jurnal Ilmiah Manajemen Informatika Dan Komputer*, 4(1), 21–27. <https://doi.org/10.32485/kopertip.v4i1.114>
- ARDIYANSYAH, RAHAYUNINGSIH, P. A., &

- MAULANA, R. 2018. Analisis Perbandingan Algoritma Klasifikasi Data Mining Untuk Dataset Blogger Dengan Rapid Miner. *Jurnal Khatulistiwa Informatika*, VI(1), 20–28.
- 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>
- HARIANDJA, T. R., & BUDIMAN, N. T. 2021. Transparansi Dalam Pelaksanaan Bantuan Langsung Tunai (Blt) Dana Desa. *Ijil*, 1(3), 263–277. <https://doi.org/10.35719/ijl.v1i3.86>
- HARIMURTI, F. A., & RIKSAKOMARA, E. 2017. *Klasifikasi Penerimaan Beasiswa Menggunakan Metode Naïve Bayes Classifier (Studi Kasus Universitas Trunojoyo Madura)*.
- IRMAYANSYAH, & FIRDAUS, A. A. 2019. Penerapan Algoritma C4.5 Untuk Klasifikasi Penentuan Penerimaan Bantuan Langsung Di Desa Ciomas. *Teknois: Jurnal Ilmiah Teknologi Informasi Dan Sains*, 8(1), 17–28. <https://doi.org/10.36350/jbs.v8i1.18>
- Kuangan, M. 2021. *Peraturan Menteri Keuangan Republik Indonesia*.
- KURNIADI, D., LESLIE, H., SPITS, H., & SUPARTA, W. 2021. Predicting Student Performance with Multi-Level Representation in an Intelligent Academic Recommender System using Backpropagation Neural Network. *ICIC Express Letters Part B: Applications*, 12(10), 883–890. <https://doi.org/10.24507/icicelb.12.10.883>
- KURNIADI, D., NURAENI, F., & JAELANI, D. (n.d.). *Implementasi Logika Fuzzy Mamdani Pada Sistem Prediksi Calon Penerima Program Keluarga Harapan*. 160–171.
- LEFYA, A. R., SAINS, F., & TEKNOLOGI, D. A. N. 2022. *Sistem Uji Kelayakan Penerima Bantuan Langsung Tunai Badan Penyelenggara Jaminan Sosial Menggunakan Metode Klasifikasi Algoritma Naïve Bayes Pada Desa Tamatto Kecamatan Ujung Loe Kabupaten Bulukumba*.
- NUL, L., IIN, H., ADE, S., & RINALDI, R. 2021. *Penerapan Data Mining Algoritma Naïve Bayes dalam Menentukan Program Keluarga Pra Sejahtera*. 1(1), 21–25.
- NURAENI, F., AGUSTIN, Y. H., RAHAYU, S., KURNIADI, D., SEPTIANA, Y., & LESTARI, S. M. 2021. Student Study Timeline Prediction Model Using Naïve Bayes Based Forward Selection Feature. *8th International Conference on ICT for Smart Society: Digital Twin for Smart Society, ICISS 2021 - Proceeding*, 1–5. <https://doi.org/10.1109/ICISS53185.2021.9532502>
- PUTRI, H., PURNAMASARI, A. I., DIKANANDA, A. R., NURDIWAN, O., & ANWAR, S. 2021. Penerima Manfaat Bantuan Non Tunai Kartu Keluarga Sejahtera Menggunakan Metode Naive Bayes dan KNN. *Building of Informatics, Technology and Science (BITS)*, 3(3), 331–337. <https://doi.org/10.47065/bits.v3i3.1093>
- SINAGA, R. W., WINANJAYA, R., & SUSIANTI, S. 2021. Analisis Data Mining Menentukan Penerima Bantuan Langsung Tunai pada Desa Pematang Purba dengan Algoritma C 4.5. *Brahmana: Jurnal Penerapan Kecerdasan Buatan*, 3(1), 1–9. <https://doi.org/10.30645/brahmana.v3i1.86>
- SIRINGORINGO, R. 2018. Klasifikasi Data Tidak Seimbang Menggunakan Algoritma SMOTE dan k-Nearest Neighbor. *Jurnal ISD*, 3(1), 44–49.
- WINTANA, D., HIKMATULLOH, H., ICHSAN, N., PURNAMA, J. J., & RAHMAWATI, A. 2019. Klasifikasi Penentuan Penerima Manfaat Program Keluarga Harapan (Pkh) Menggunakan Algoritma C5.0 (Studi Kasus: Desa Sukamaju, Kec.Kadudampit). *Klik - Kumpulan Jurnal Ilmu Komputer*, 6(3), 254. <https://doi.org/10.20527/klik.v6i3.206>

Halaman ini sengaja dikosongkan