



## PENERAPAN *NAÏVE BAYES* DALAM MENGLASIFIKASIKAN MASYARAKAT MISKIN DI DESA LEPAK

Wiwit Pura Nurmayanti<sup>\*1</sup>, Dinda Ayu Lara Saky<sup>2</sup>, Muhammad Malthuf<sup>3</sup>,  
Muhammad Gazali<sup>4</sup>, Ristu Haiban Hirzi<sup>5</sup>,

<sup>1,2,4,5</sup> Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Hamzanwadi, Selong, Indonesia

<sup>3</sup> Program Studi Tadris IPS Fakultas Tarbiyah dan Keguruan Universitas Islam Negeri Mataram, Mataram, Indonesia

\*Email Koresponden: [wiwit.adiwinata3@gmail.com](mailto:wiwit.adiwinata3@gmail.com)

Diterima: 09-05-2021, Revisi: 20-05-2021, Disetujui: 07-06-2021

©2021 Program Studi Pendidikan Geografi, FISE, Universitas Hamzanwadi

**Abstrak** Desa Lepak merupakan desa yang berada di pulau Lombok Nusa Tenggara Barat dengan angka penduduk miskin cukup tinggi. Penelitian ini bertujuan untuk mengklasifikasi masyarakat di desa Lepak Kecamatan Sakra Timur Kabupaten Lombok Timur. Teknik pengumpulan data dilakukan dengan teknik studi dokumen. Teknik analisis data pada penelitian ini menggunakan metode *Naïve Bayes Classifier*, yang merupakan salah satu teknik pengklasifikasian dalam data mining. Berdasarkan hasil pengujian *confusion matrix* diperoleh klasifikasi masyarakat miskin di desa Lepak yang memang miskin adalah 148 *record* dari 156 *record* yang artinya terdapat 8 *record* yang *error* dimana ia lebih mirip dengan yang tidak miskin. Sedangkan untuk masyarakat tidak miskin terdapat 110 *record* dari 111 *record* yang memang tidak miskin dan sisanya 1 *record error* yang lebih mirip dengan miskin. Keakuratan data *testing* dalam memprediksi hasil klasifikasi yang menunjukkan masyarakat miskin dan tidak miskin dapat dilihat dari nilai *accuracy* yaitu sebesar 96.63% yang artinya termasuk dalam kategori *good*. Berdasarkan penelitian ini menunjukkan bahwa klasifikasi kelas untuk masyarakat desa Lepak adalah kelas dengan masyarakat miskin.

**Kata kunci:** Tingkat kemiskinan, Data Mining, Klasifikasi, *Naïve Bayes*

**Abstract** *Lepak Village is located on the Lombok island, West Nusa Tenggara with a fairly high number of poor people. This study aims to classify the community in Lepak Village, Sakra Timur District, East Lombok Regency. The data collection technique is done by document study technique. The data analysis technique in this study uses the Naïve Bayes Classifier method, which is one of the classification techniques in data mining. Based on the results of confusion matrix testing, it was found that the classification of the poor in Lepak village who are poor is 148 records out of 156 records, which means that there are 8 error records where they are more similar to those who are not poor. Meanwhile, for the non-poor, there were 110 out of 111 records that were not poor and the remaining 1 record error was more similar to poor. The accuracy of the testing data in predicting the classification results which shows that the poor and not poor can be seen from the accuracy value, which is 96.63%, which means it is included in the good category. Based on this, it can be stated that the class classification for the Lepak village community is the class with poor people.*

**Keywords:** *Poverty rate, Data Mining, Classification, Naïve Bayes*

## PENDAHULUAN

Masalah kemiskinan selalu memperoleh perhatian utama di Indonesia. Hal ini akan menyebabkan munculnya masalah baru dari berbagai persoalan sosial, ekonomi dan politik ditengah-tengah masyarakat (Purwanto, 2004). Kemiskinan sendiri menunjukkan bahwa seseorang yang tidak mampu lagi memenuhi kebutuhan dasar baik pangan, tempat tinggal, sandang, pendidikan dan kesehatan secara layak dimana penetapan garis kemiskinan berdasarkan tingkat perekonomian (Abdi, 2019). Hal ini mengakibatkan suatu daerah dikategorikan tingkat kemiskinannya tinggi jika persoalan tersebut tidak bisa terselesaikan. Berdasarkan data yang dirilis oleh Badan Pusat Statistik (BPS) pada Maret 2020 bahwa jumlah penduduk miskin mencapai 26,42 juta orang dibandingkan September 2019, jumlah

penduduk miskin meningkat 1,63 juta orang, sementara jika dibandingkan dengan Maret 2019, jumlah penduduk miskin meningkat sebanyak 1,28 juta orang. Persentase penduduk miskin pada Maret 2020 tercatat sebesar 9,78 persen, meningkat 0,56 persen poin terhadap September 2019 dan meningkat 0,37 persen poin terhadap Maret 2019 (Badan Pusat Statistik, 2020).

**Tabel 1.** Persentase dan Jumlah Penduduk Miskin menurut Pulau, Maret 2020

Pulau	Persentase Penduduk Miskin			Jumlah Penduduk Miskin (ribu orang)		
	Perkotaan	Perdesaan	Total	Perkotaan	Perdesaan	Total
Sumatera	8,39	11,01	9,87	2.154,41	3.683,06	5.837,47
Jawa	7,48	12,64	9,24	7.493,53	6.559,28	14.052,81
Bali dan Nusa Tenggara	8,56	17,69	13,55	582,20	1.450,63	2.032,83
Kalimantan	4,35	7,13	5,81	345,36	624,28	969,64
Sulawesi	5,63	13,12	10,10	452,28	1.554,50	2.006,78
Maluku dan Papua	5,25	28,15	20,34	134,18	1.390,31	1.524,49
Indonesia	7,38	12,82	9,78	11.161,96	15.262,06	26.424,02

Sumber: Diolah dari data survei sosial ekonomi nasional (Susenas) Maret 2020

Data yang ditampilkan pada Tabel 1 menunjukkan bahwa persentase penduduk miskin daerah Nusa Tenggara khususnya wilayah perkotaan sebesar 8,56% dan ini menunjukkan bahwa angka tersebut merupakan angka tertinggi dibandingkan dengan pulau lainnya dan persentase penduduk miskin di perdesaan sendiri sebesar 17,69%, angka ini merupakan angka kedua tertinggi setelah pulau Maluku dan Papua dengan jumlah penduduk miskin di wilayah pedesaan sebesar 1.450,63 ribu orang. Berdasarkan UU Nomor 64 tahun 1958, Nusa Tenggara Barat (NTB) merupakan salah satu dari tiga Daerah Tingkat I yang ada di Provinsi Nusa Tenggara.

NTB memiliki kondisi alam yang sering mengalami kekeringan dan rawan pangan sehingga mengakibatkan gizi buruk, dengan akses fasilitas kesehatan yang kurang memadai menjadi persoalan tersendiri, selain itu angka putus sekolah juga cukup tinggi, permasalahan lain juga masih menjadi perhatian yaitu tingkat kemiskinan. Berdasarkan data yang dirilis BPS September 2020, NTB masih masuk dalam 10 provinsi dengan persentase penduduk miskin tertinggi. Masih tingginya angka kemiskinan di setiap Kabupaten/Kota di Provinsi NTB membuat provinsi ini terus dilanda permasalahan kemiskinan, dengan tingginya kemiskinan maka penanganan yang serius dari pemerintah maupun stakeholders sangat diperlukan guna meminimalisir angka kemiskinan tersebut. Sejak tahun 2001 diberlakukannya kebijakan otonomi daerah dimana pemerintah daerah berwenang penuh merancang dan melaksanakan kebijakan dan program pembangunan sesuai dengan kebutuhannya. Dalam pelaksanaan otonomi daerah, pemerintah daerah tidak hanya melaksanakan program pembangunan tetapi juga bertanggung jawab secara langsung dan aktif dalam penanganan kemiskinan, sehingga untuk menanggulangi kemiskinan perlu dibutuhkan informasi terkait daerah mana saja yang masuk dalam kategori miskin.

Dalam penelitian Annur (2018) mengenai “Klasifikasi Masyarakat Miskin Menggunakan Metode *Naïve Bayes*” dengan studi kasus Kecamatan Tibawa Kabupaten Gorontalo dihasilkan kesimpulan bahwa, sistem klasifikasi masyarakat miskin di wilayah pemerintahan Tibawa dapat direkayasa dan berdasarkan hasil pengujian *confusion matrix* dengan teknik split validasi, penggunaan metode klasifikasi *Naïve Bayes* terhadap dataset yang telah diambil pada objek penelitian diperoleh tingkat akurasi sebesar 73% atau termasuk dalam kategori *Good*. Sementara nilai *Precision* sebesar 92% dan *Recall* sebesar 86% (Annur, 2018). Kemudian dalam penelitian Simatupang, Wuryandari, & Suparti (2016) tentang “Klasifikasi Rumah Layak Huni di Kabupaten Brebes dengan Menggunakan Metode *Learning Vector Quantization* dan *Naïve Bayes*” menunjukkan hasil klasifikasi rumah layak huni dengan menggunakan metode *Learning Vector Quantization* adalah 30 rumah tepat klasifikasi dan 12 rumah tidak tepat klasifikasi. Ketepatan klasifikasi terbaik pada metode *Learning Vector Quantization* dengan proporsi pembagian data 95% data latih dan 5% data uji sebesar 71,43% dan laju *error* sebesar 28,57%. Hasil klasifikasi rumah layak huni dengan menggunakan metode *Naïve Bayes Classifier* adalah 40 rumah tepat

klasifikasi dan 2 rumah tidak tepat klasifikasi. Ketepatan klasifikasi terbaik pada metode *Naive Bayes Classifier* dengan proporsi pembagian data 95% data latih dan 5% data uji sebesar 95,24% dan laju error sebesar 4,76%. Berdasarkan nilai akurasi yang diperoleh, metode *Naive Bayes Classifier* menghasilkan nilai akurasi lebih tinggi dibandingkan metode *Learning Vector Quantization*. Sehingga metode *Naive Bayes Classifier* lebih baik dalam mengklasifikasikan rumah layak huni di Kabupaten Brebes (Simatupang, Wuryandari, & Suparti, 2016; Mujahid & Marsoyo, 2019).

Desa Lepak merupakan salah satu desa yang berada di kabupaten Lombok Timur NTB, dimana Lombok Timur sendiri berada di urutan ke sembilan dari 10 kabupaten/kota penduduk miskin di NTB (Badan Pusat Statistik, 2020). Data penduduk miskin desa Lepak belum dimanfaatkan untuk penelitian dan sebatas arsip untuk pemerintah desa, sehingga peneliti ingin memanfaatkan data tersebut untuk dilakukan analisis klasifikasi guna memperoleh informasi menggunakan teknik data mining, yaitu suatu proses pengumpulan informasi penting dari suatu data yang besar, dimana dalam prosesnya seringkali menggunakan metode statistika dan matematika (Santosa, 2007). *Output* dari analisis ini diharapkan dapat memberikan informasi yang bisa membantu pihak desa untuk merancang strategi dalam meningkatkan kesejahteraan masyarakat. Variabel yang digunakan dalam melakukan klasifikasi penduduk miskin adalah Jenis Kelamin, Peserta PKH, Jumlah Anggota Rumah Tangga, Status kepemilikan rumah, Fasilitas rumah, Luas lantai rumah, Fasilitas elektronik, sesuai data yang telah diambil dan sesuai dengan variabel yang diinputkan, maka hasil klasifikasinya dapat menentukan apakah masyarakat desa Lepak termasuk kedalam kelas miskin dan tidak miskin.

## METODE PENELITIAN

### Data, Variabel dan Definisi Operasional Variabel

Data yang digunakan dalam penelitian ini adalah data sekunder terkait penduduk desa Lepak yang diperoleh dari pemerintah desa Lepak kecamatan Sakra Timur, Kabupaten Lombok Timur, NTB. Adapun data yang disediakan pihak desa berupa data variabel jenis kelamin, peserta PKH, jumlah anggota rumah tangga, status kepemilikan rumah, fasilitas rumah, luas lantai rumah, dan fasilitas elektronik. Adapun variabel dan definisi operasional variabel penelitian yaitu: 1) Status kepemilikan rumah ( $X_1$ ), rumah yang ditempati masyarakat desa Lepak memiliki beberapa status kepemilikan yaitu, diantaranya rumah milik sendiri, rumah bebas sewa, kontrak dan lainnya. Jadi status kepemilikan rumah yang dimaksud adalah apakah rumah yang ditempati merupakan milik sendiri atau lainnya; 2) Jenis Kelamin ( $X_2$ ), jenis kelamin yang dimaksud adalah jenis kelamin dari kepala keluarga, apakah kepala keluarga dari suatu rumah tangga tersebut laki-laki atau perempuan; 3) Peserta PKH ( $X_3$ ), Program Keluarga Harapan (PKH) adalah program pemberian bantuan sosial bersyarat kepada keluarga miskin yang ditetapkan sebagai keluarga penerima manfaat PKH; 4) Fasilitas rumah ( $X_4$ ), apakah keluarga tersebut memiliki fasilitas rumah yang lengkap atau tidak, seperti memiliki jamban dan lain-lain; 5) Fasilitas elektronik ( $X_5$ ), apakah dalam keluarga tersebut memiliki fasilitas elektronik salah satunya seperti memiliki television atau tidak; 6) Jumlah anggota keluarga ( $X_6$ ), Banyaknya anggota rumah tangga dalam satu kartu keluarga. Misalnya, terdapat 5 orang anggota keluarga (termasuk kepala keluarga) dalam satu kartu keluarga; 7) Luas lantai rumah ( $X_7$ ), Luas lantai rumah yang dimaksud adalah luas lantai tempat tinggal yang ditempati oleh keluarga tersebut. Misalnya, luas lantai rumah yang ditempati keluarga A adalah 42 m<sup>2</sup>.

### Klasifikasi

Klasifikasi merupakan proses penemuan model (atau fungsi) yang menggambarkan dan membedakan kelas data atau konsep yang bertujuan agar bisa digunakan untuk memprediksi kelas dari objek yang label kelasnya tidak diketahui (Kusrini, 2009). Algoritma klasifikasi yang banyak digunakan secara luas, yaitu *Decision/classification trees*, *Bayesian classifiers/Naive Bayes classifiers*, *Neural networks*, Analisa Statistik, Algoritma Genetika, *Rough sets*, *K-nearest neighbor*, *Metode Rule Based*, *Memory based reasoning*, dan *Support vector machines (SVM)* (Jananto, 2013). Untuk mengevaluasi kinerja suatu metode klasifikasi dilakukan dengan *confusion matrix*. *Confusion matrix* dapat dilihat pada Tabel 2.

Tabel 2. Confusion matrix kelas biner

Class	Predictive positive	Predictive negative
Actual positive	True positive (TP)	False negative (FN)
Actual negative	False positive (FP)	True Negative (TN)

Untuk menghitung akurasi dan laju *error* digunakan formula (Gorunescu, 2011):

$$Accuracy = \frac{TP+TN}{Total} \dots\dots\dots(1)$$

$$Error = \frac{FP+FN}{Total} \dots\dots\dots(2)$$

**Naïve Bayes Clasifier**

Algoritma *Naive Bayes* adalah salah satu algoritma yang terdapat pada teknik data mining klasifikasi. *Naive Bayes* merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuan Inggris yaitu Thomas Bayes, *Naïve Bayes* memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya, sehingga dikenal dengan Teorema Bayes. Teorema tersebut dikombinasikan dengan naïve dimana diasumsikan kondisi antar atribut saling bebas. Klasifikasi *Naïve Bayes* diasumsikan bahwa ada atau tidak ciri tertentu dari sebuah kelas tidak ada hubungannya dengan ciri dari kelas lainnya (Bustami, 2013).

Dalam menggunakan metode *Naive Bayes*, keuntungan yang bisa didapat adalah metode ini hanya membutuhkan jumlah data training yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian. Karena diasumsikan sebagai variabel independen, maka hanya varian dari suatu variabel dalam sebuah kelas yang dibutuhkan untuk menentukan klasifikasi, bukan keseluruhan dari matriks kovarians (Santoso, 2017). *Naive Bayes* merupakan teknik prediksi berbasis probabilitas sederhana yang berdasarkan pada penerapan teorema *Bayes* dengan asumsi independensi yang kuat (Saleh, 2015). *Naive Bayes* didasarkan pada asumsi penyederhanaan bahwa nilai atribut secara kondisional saling bebas jika diberikan nilai *output* (Santoso, 2017). Pengklasifikasian *Naive Bayes* dilakukan dengan memilih probabilitas akhir (posterior) tertinggi dari masing-masing kelas (Simatupang, Wuryandari, & Suparti, 2016).

Teorema Bayes memiliki bentuk umum sebagai berikut (Prasetyo, 2012):

$$P(Y|X_1, X_2, \dots, X_p) = \frac{P(Y)P(X_1, X_2, \dots, X_p|Y)}{P(X_1, X_2, \dots, X_p)} \dots\dots\dots(3)$$

dengan:

- $P(Y|X_1, X_2, \dots, X_p)$  = probabilitas masuknya obyek dengan karakteristik variabel tertentu dalam kelompok Y (*posterior*)
- $P(X_1, X_2, \dots, X_p|Y)$  = probabilitas kemunculan variabel-variabel pada obyek yang masuk kelompok Y (*likelihood*)
- $P(Y)$  = probabilitas munculnya kelompok Y sebelum masuknya obyek (*prior*)
- $P(X_1, X_2, \dots, X_p)$  = peluang kemunculan variabel-variabel pada obyek secara umum (*evidence*).

Pada persamaan (3) dapat juga dituliskan secara sederhana sebagai berikut:

$$Posterior = \frac{Prior \times Likelihood}{Evidence} \dots\dots\dots(4)$$

Nilai dari posterior tersebut dibandingkan dengan nilai-nilai posterior kelompok lainnya untuk menentukan kelompok suatu obyek yang diklasifikasikan. Mengklasifikasikan suatu obyek dapat ditentukan dengan memilih kelompok yang memiliki posterior terbesar, nilai *evidence* selalu tetap untuk setiap kelompok pada satu sampel yaitu bernilai 1 dan merupakan pembagi pada setiap kelompok sehingga dalam perhitungan posterior hanya cukup mengalikan nilai prior dengan likelihood. Nilai prior yang merupakan peluang munculnya kelompok Y sebelum masuknya obyek dapat dihitung menggunakan persamaan sebagai berikut:

$$P(Y_g) \frac{n_g}{N} \dots\dots\dots(5)$$

dengan:

$P(Y_g)$  = peluang munculnya kelompok Y ke- $g$ ,  $g = 1, 2, \dots, q$

$n_g$  = banyaknya pengamatan pada kelompok ke- $g$

Penjabaran lebih lanjut rumus Bayes tersebut dilakukan dengan menjabarkan  $P(Y|X_1, X_2, \dots, X_p)$  menggunakan aturan perkalian sebagai berikut:

$$\begin{aligned} P(Y|X_1, X_2, \dots, X_p) &= P(Y)P(X_1|Y)P(X_2, X_3, \dots, X_p|Y, X_1) \\ &= P(Y)P(X_1|Y)P(X_2, X_3, \dots, X_p|Y, X_1)P(X_3, X_4, \dots, X_p|Y, X_1, X_2) \dots\dots\dots(6) \\ &= P(Y)P(X_1|Y)P(X_2|Y, X_1)P(X_3|Y, X_1, X_2) \dots \\ &\quad P(X_4, X_5, \dots, X_p|Y, X_1, X_2, X_3) \\ &= P(Y)P(X_1|Y)P(X_2|Y, X_1)P(X_3|Y, X_1, X_2) \dots \\ &\quad P(X_p|Y, X_1, X_2, X_3, \dots, X_{p-1}) \end{aligned}$$

Dapat dilihat bahwa hasil penjabaran tersebut menyebabkan semakin banyak dan semakin kompleksnya faktor-faktor syarat yang mempengaruhi nilai peluang yang hampir mustahil untuk dianalisa satu persatu. Kompleksnya faktor-faktor syarat yang mempengaruhi nilai peluang menyebabkan perhitungan tersebut menjadi sulit untuk dilakukan, maka digunakan asumsi independensi yang sangat tinggi (*naive*), bahwa masing-masing petunjuk ( $X_1, X_2, \dots, X_p$ ) saling bebas (*independent*) satu sama lain sehingga berlaku suatu kesamaan sebagai berikut:

$$P(X_a|X_b) = \frac{P(X_a \cap X_b)}{P(X_b)} = \frac{P(X_a)P(X_b)}{P(X_b)} = P(X_a) \dots\dots\dots(7)$$

untuk  $a \neq b$  sehingga  $P(X_a|Y, X_b) = P(X_a|Y)$

Pada persamaan (7) dapat disimpulkan bahwa asumsi independensi *naive* tersebut membuat syarat peluang menjadi sederhana sehingga perhitungan menjadi mungkin dilakukan. Selanjutnya penjabaran  $P(Y|X_1, X_2, \dots, X_p)$  dapat disederhanakan menjadi:

$$\begin{aligned} P(Y|X_1, X_2, \dots, X_p) &= P(Y)P(X_1|Y)P(X_2|Y)P(X_3|Y) \dots P(X_p|Y) \\ &= P(Y) \prod_{k=1}^p P(X_k|Y) \dots\dots\dots(8) \end{aligned}$$

Persamaan (8) merupakan model dari teorema *Naive Bayes* yang selanjutnya digunakan dalam proses klasifikasi.

### Tahapan Analisis

Pada umumnya teorema *Naive Bayes* mudah dihitung untuk nilai pengamatan variabel bebas bertipe kategorik, namun untuk nilai pengamatan variabel bebas dengan tipe numerik (*non* kategorik) ada perlakuan khusus sebelum diproses menggunakan *Naive Bayes* yaitu dengan cara sebagai berikut: 1) Melakukan diskritisasi pada setiap nilai pengamatan variabel bebas kontinu dan mengganti nilai pengamatan tersebut dengan nilai interval diskrit. Pendekatan ini dilakukan dengan mentransformasi ke

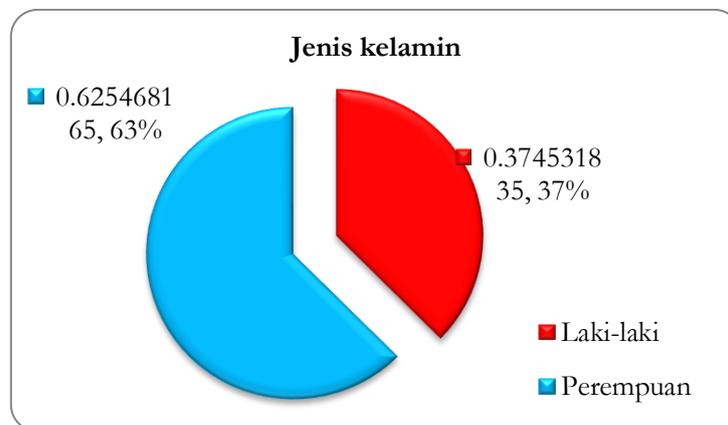
dalam skala ordinal; 2) Mengasumsikan bentuk tertentu dari distribusi peluang untuk nilai pengamatan kontinu dan memperkirakan parameter distribusi dengan data training.

Berikut adalah tahapan analisis dalam penelitian ini untuk mengklasifikasikan dengan metode *Naïve Bayes* menggunakan program *R-Studio*: 1) Menginput dataset; 2) Membagi data menjadi 2, yaitu data *train* dan data *test* dengan sekali percobaan proporsi pembagian data yaitu 75% data *train* dan 25% data *test*; 3) Membuat model dengan data *training*; 4) Memprediksi model dengan data *testing*; dan 5) Terakhir melihat performa kinerja klasifikasi dengan *confusion matrix*.

Beberapa indikator kemiskinan pada satu rumah tangga yang ditentukan adalah: 1) Luas lantai bangunan tempat tinggal kurang dari 8 m<sup>2</sup> per orang; 2) Jenis lantai bangunan tempat tinggal terbuat dari tanah/bambu/kayu murahan; 3) Jenis dinding tempat tinggal terbuat dari bambu/rumbia/kayu berkualitas rendah/tembok tanpa plester; 4) Tidak memiliki fasilitas buang air besar atau bersama-sama dengan rumah tangga lain; 5) Sumber penerangan rumah tangga tidak menggunakan listrik; 6) Sumber air minum berasal dari sumur/mata air tidak terlindungi/sungai/ air hujan; 7) Bahan bakar untuk memasak sehari-hari adalah kayu bakar/arang/ minyak tanah; 8) Hanya mengkonsumsi daging/susu/ayam satu kali dalam seminggu; 9) Hanya membeli satu stel pakaian baru dalam setahun; 10) Hanya sanggup makan sebanyak satu/dua kali dalam sehari; 11) Tidak sanggup membayar biaya pengobatan di Puskesmas/Poliklinik; 12) Sumber penghasilan kepala rumah tangga adalah petani dengan luas lahan 0,5 ha, buruh tani, nelayan, buruh perkebunan atau pekerjaan lainnya dengan pendapatan di bawah Rp. 600.000,- (Enam ratus ribu rupiah); 13) Pendidikan tertinggi kepala rumah tangga adalah tidak sekolah/tidak tamat SD/hanya SD; dan 14) Tidak memiliki tabungan/barang yang mudah dijual dengan nilai Rp. 500.000,- (Lima ratus ribu rupiah) seperti sepeda motor (kredit/non kredit), emas, ternak, kapal motor atau barang modal lainnya.

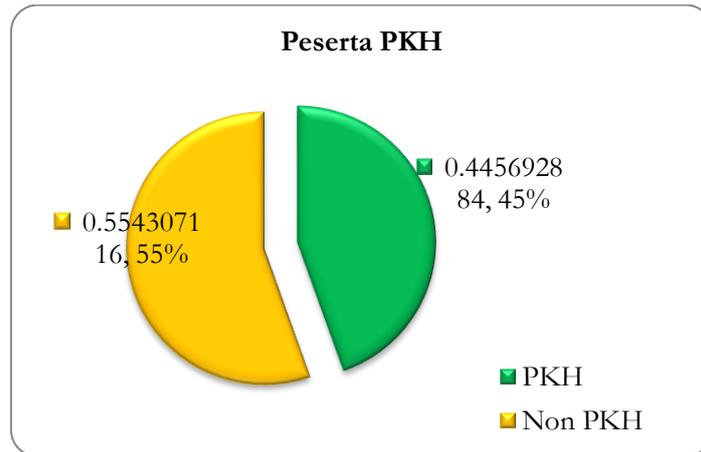
## TEMUAN DAN PEMBAHASAN

Pada proses pengukuran ini dilakukan menggunakan analisis deskriptif dan analisis *Naïve Bayes Classifier* dan menganalisis data menggunakan *excel* dan *R-studio*. Berikut adalah sajian analisis deskriptif berdasarkan data yang digunakan:



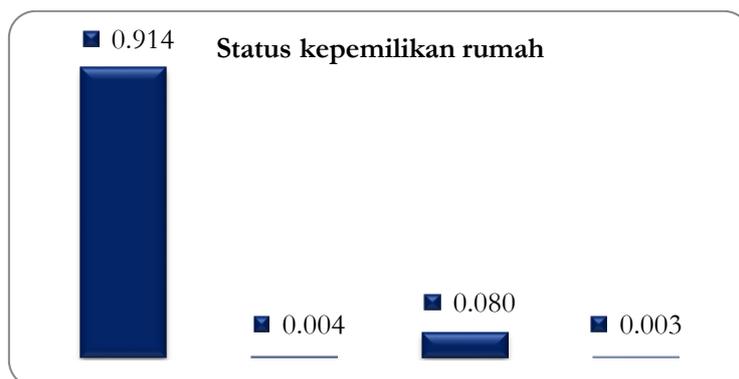
**Gambar 1.** Persentase jenis kelamin kepala keluarga  
(Sumber: Hasil analisis peneliti, 2021)

Bersarkan Gambar 1 dapat diketahui bahwa persentase jenis kelamin kepala keluarga di desa Lepak mayoritas adalah perempuan yaitu sebesar 63% dan hanya 37% laki-laki yang menjadi kepala keluarga di desa Lepak. Ada beberapa alasan perempuan menjadi kepala keluarga seperti suami meninggal, suami pengangguran atau sakit, bercerai dan dipoligami. Keluarga yang dikepalai perempuan merupakan keluarga yang paling rentan terhadap masalah ekonomi. Berdasarkan informasi dari BPS bahwa ada 67,17% perempuan menjadi kepala rumah tangga dikarenakan suaminya meninggal. Data BPS yang sama menunjukan 42,57 persen tidak punya ijazah, dan NTB merupakan salah satu provinsi yang jumlah perempuan menjadi kepala keluarga paling besar.



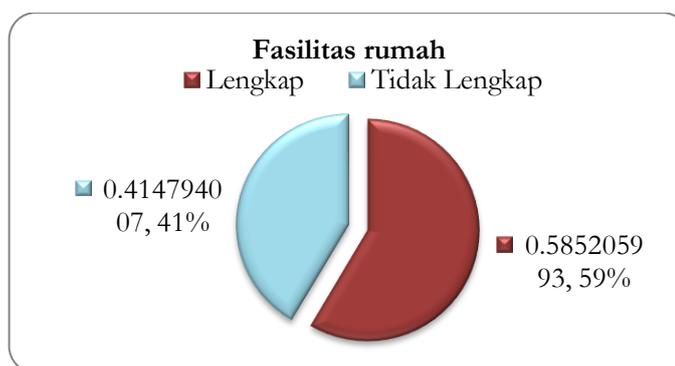
**Gambar 2.** Persentase masyarakat yang menerima PKH (Sumber: Hasil analisis peneliti, 2021)

Berdasarkan Gambar 2 dapat diketahui bahwa persentase masyarakat desa Lepak yang menerima Program Keluarga Harapan (PKH) ada sebanyak 45% dan sisanya 55% tidak menerima bantuan PKH. PKH merupakan program pemberian bantuan sosial bersyarat kepada keluarga miskin yang ditetapkan sebagai keluarga penerima manfaat. PKH sejatinya harus dijalankan dengan prinsip keadilan, tanggung jawab, dan takhful (menjamin) (Kartiawati, 2017). Desa Lepak sendiri memiliki cukup banyak warga yang menerima program pemerintah tersebut. Hal ini dapat dilihat sebagaimana ditampilkan pada gambar 3.



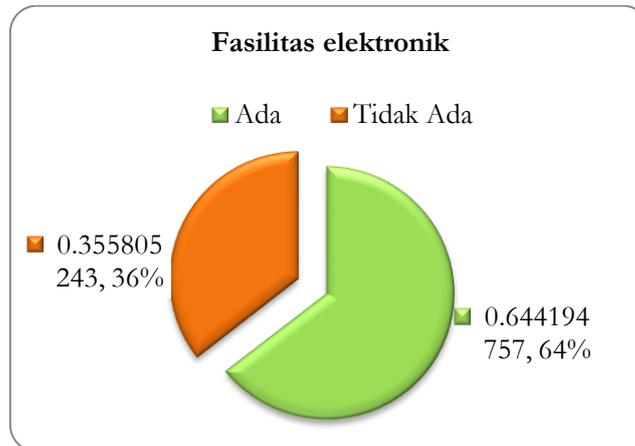
**Gambar 3.** Persentase status kepemilikan rumah (Sumber: Hasil analisis peneliti, 2021)

Berdasarkan Gambar 3 dapat diketahui bahwa persentase status kepemilikan rumah masyarakat desa Lepak adalah terdapat 91.4% atau 91.4% merupakan rumah milik sendiri, 8% merupakan rumah bebas sewa dan status kepemilikan rumah kontrak dan lainnya memiliki persentase 0 dibandingkan yang lain.



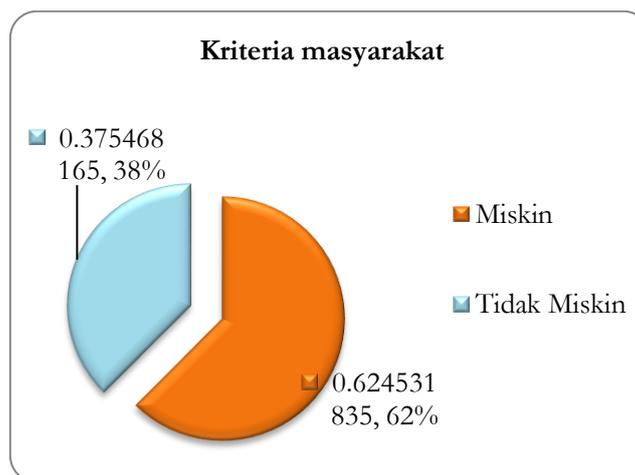
**Gambar 4.** Persentase fasilitas rumah (Sumber: Hasil analisis peneliti, 2021)

Berdasarkan Gambar 4 dapat diketahui bahwa persentase fasilitas rumah masyarakat desa Lepak adalah terdapat 41% dikategorikan fasilitas rumah yang lengkap dan 59% dengan kategori fasilitas rumah yang tidak lengkap.



**Gambar 5.** Persentase fasilitas elektronik  
(Sumber: Hasil analisis peneliti, 2021)

Berdasarkan Gambar 5 dapat diketahui bahwa persentase fasilitas elektronik yang dimiliki masyarakat desa Lepak adalah terdapat 36% dikategorikan ada dan sisanya 64% tergolong tidak ada.



**Gambar 6.** Persentase kriteria masyarakat  
(Sumber: Hasil analisis peneliti, 2021)

Berdasarkan Gambar 6 dapat diketahui bahwa persentase kriteria masyarakat desa Lepak adalah 38% tergolong masyarakat tidak miskin dan 62% tergolong masyarakat miskin.

Berdasarkan Gambar 1 s/d Gambar 5 tersebut dapat diketahui bahwa masyarakat di desa Lepak mayoritas merupakan masyarakat miskin dengan jenis kelamin kepala keluarga perempuan yang bukan merupakan peserta PKH dan rumah yang ditempati merupakan rumah milik sendiri serta fasilitas rumah baik seperti memiliki jamban dan *television* masyarakat desa Lepak memilikinya. Sementara berdasarkan Gambar 6 dapat diketahui bahwa data kriteria masyarakat miskin tersebut masih tergolong seimbang karena < 80:20, sehingga metode *single* klasifikasi *Naïve Bayes* bisa digunakan.

Hasil proses klasifikasi menggunakan metode *Naïve Bayes* dimana dataset yang digunakan dibagi kedalam dua bagian yakni 75% (801 *record*) dari dataset dijadikan sebagai data *training* dan 25% (267 *record*) sisanya dijadikan sebagai data *testing*. Dengan menggunakan data *training* sebanyak 801 *record* untuk membuat model dan memprediksi model tersebut dengan data *testing* sebanyak 207 *recorde*, sehingga dengan hasil prediksi tersebut didapat klasifikasi data dan performa kinerja klasifikasi tersebut. Berikut adalah tabel *confusion matrix*:

Tabel 3. Performa *confusion matrix*

Confusion Matrix and Statistics		
Prediction	Reference	
	Miskin	Tidak Miskin
Miskin	148	1
Tidak Miskin	8	110
Accuracy : 0.9663		
95% CI : (0.937, 0.9845)		
No Information Rate : 0.5843		
P-Value [Acc > NIR] : <2e-16		
Kappa : 0.9312		
McNemar's Test P-Value : 0.0455		
Sensitivity : 0.9487		
Specificity : 0.9910		
Pos Pred Value : 0.9933		
Neg Pred Value : 0.9322		
Prevalence : 0.5843		
Detection Rate : 0.5543		
Detection Prevalence : 0.5581		
Balanced Accuracy : 0.9699		
'Positive' Class : Miskin		

(Sumber: Hasil analisis peneliti, 2021)

Tabel 4. Output klasifikasi masyarakat miskin

Class		Predictive	
		positive	negative
Actual	positive	148	1
	negative	8	110

Sumber: Hasil analisis peneliti, 2021.

Berdasarkan Tabel 3 dan Tabel 4 performa *confusion matrix* dapat diketahui bahwa prediksi masyarakat miskin yang memang miskin adalah 148 *record* dari 156 *record* yang artinya terdapat 8 *record* yang *error*, dimana ia lebih mirip dengan yang tidak miskin. Sedangkan untuk prediksi masyarakat tidak miskin terdapat 110 *record* dari 111 *record* yang memang tidak miskin dan sisanya 1 *record* yang *error*, dimana ia lebih mirip dengan yang miskin. Dilihat dari *positive class* dapat diketahui bahwa klasifikasi kelas untuk masyarakat desa Lepak adalah kelas miskin.

Untuk melihat keakuratan data *testing* dalam memprediksi hasil klasifikasi yang menunjukkan masyarakat miskin dan tidak miskin dapat dilihat dari nilai *accuracy* yaitu sebesar 96.63% yang artinya termasuk dalam kategori *good* dalam memprediksi. Dan keakuratan tes dicerminkan dengan *sensitivity* dan *specificity*. Dari Tabel 3 tersebut dapat diketahui bahwa tingkat *accuracy* sebesar 96.63% yang artinya termasuk dalam kategori *good* dalam memprediksi. Sementara nilai *sensitivity* yang menunjukkan prediksi untuk kelas miskin sebesar 94.87% dan *specificity* yang menunjukkan prediksi untuk kelas tidak miskin sebesar 99.1%. Berdasarkan hal tersebut dapat dinyatakan bahwa sistem klasifikasi yang dibangun dapat digunakan sebagai bahan masukan bagi pengambil keputusan.

## SIMPULAN

Berdasarkan hasil pengujian *confusion matrix* dengan penggunaan metode klasifikasi *Naïve Bayes* terhadap dataset yang telah diambil pada objek penelitian diperoleh klasifikasi masyarakat miskin di desa Lepak yang memang miskin adalah 148 *record* dari 156 *record* yang artinya terdapat 8 *record* yang *error*, dimana ia lebih mirip dengan yang tidak miskin. Sedangkan untuk klasifikasi masyarakat tidak miskin terdapat 110 *record* dari 111 *record* yang memang tidak miskin dan sisanya 1 *record* yang *error*, dimana ia lebih mirip dengan yang miskin. Dilihat dari *positive class* dapat diketahui bahwa klasifikasi kelas untuk masyarakat desa Lepak adalah kelas dengan masyarakat miskin.

## DAFTAR PUSTAKA

- Abdi, H. (2019). 5 Penyebab Kemiskinan dan Definisinya yang Wajib Diketahui. *Http://Www.Liputan6.Com*, p. 1. Retrieved from <https://hot.liputan6.com/read/3936545/5-penyebab-kemiskinan-dan-definisinya-yang-wajib-diketahui>.
- Annur, H. (2018). Klasifikasi Masyarakat Miskin Menggunakan Metode Naive Bayes. *ILKOM Jurnal Ilmiah*, 10(2), 160-165.
- Badan Pusat Statistik. (2020). Profil Kemiskinan di Indonesia Maret 2020. Retrieved July 15, 2020, from <http://www.bps.go.id>.
- Bustami, B. (2013). Penerapan Algoritma Naive Bayes Untuk Mengklasifikasi Data Nasabah Asuransi. *TECHSI-Jurnal Teknik Informatika*, 5(2), 127-146.
- Gorunescu, F. (2011). *Data Mining: Concepts, Models and Techniques* (Vol. 12). Springer Science & Business Media.
- Jananto, A. (2013). Algoritma Naive Bayes untuk Mencari Perkiraan Waktu Studi Mahasiswa. *Dinamik*, 18(1), 9-16.
- Kartiawati, K. (2017). Analisis Efektivitas Keluarga Harapan (PKH) Dalam Pengentasan Kemiskinan Ditinjau Dari Perspektif Ekonomi Islam (Studi Pada Peserta PKH Kampung Bonglai Kecamatan Banjit Kabupaten Way Kanan). *Skripsi*. UIN Raden Intan Lampung.
- Kusrini, L. (2009). *Algoritma Data Mining*. Yogyakarta: Andi Offset.
- Mujahid, A. S., & Marsoyo, A. (2019). Perbandingan Nilai Ekonomi Lahan dalam Kasus Konversi Lahan Sawah di Kecamatan Praya Kabupaten Lombok Tengah. *Geodika: Jurnal Kajian Ilmu dan Pendidikan Geografi*, 3(2), 58-69.
- Prasetyo, E. (2012). *Data Mining Konsep dan Aplikasi Menggunakan Matlab*. Yogyakarta: Andi.
- Purwanto, E. A. (2004). Mengkaji Potensi Usaha Kecil Dan Menengah (UKM) untuk Pembuatan Kebijakan Anti Kemiskinan di Indonesia (JSP Volume 10 No 3). *Jurnal (Mengkaji Potensi Usaha Kecil Dan Menengah (UKM) Untuk Pembuatan Kebijakan Anti Kemiskinan Di Indonesia" Penerbit FISIPOL UGM, Volume 10 No. 3, Maret 2007 ISSN: 1410-4946)*, 100(23), 295-324.
- Saleh, A. (2015). Implementasi Metode Klasifikasi Naive Bayes dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga. *Creative Information Technology Journal*, 2(3), 207-217.
- Santosa, B. (2007). *Data Mining; Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Yogyakarta: Graha Ilmu.
- Santoso, H. (2017). *Data Mining Penyusunan Buku Perpustakaan Daerah Lombok Barat Menggunakan Algoritma Apriori*. Seminar Nasional TIK dan Ilmu Sosial.
- Simatupang, F. J., Wuryandari, T., & Suparti, S. (2016). Klasifikasi Rumah Layak Huni Di Kabupaten Brebes Dengan Menggunakan Metode Learning Vector Quantization Dan Naive Bayes. *Jurnal Gaussian*, 5(1), 99-111.
- Undang-undang Republik Indonesia Nomor 64 Tahun 1958 tentang Pembentukan Daerah-daerah Tingkat I Bali, Nusa Tenggara Barat dan Nusa Tenggara Timur.