

PENENTUAN KLASIFIKASI TINGKAT KESEJAHTERAAN KELUARGA MENGGUNAKAN METODE NAIVE BAYES PADA KECAMATAN PASAR JAMBI

*Ade Purnamasari*¹, *Setiawan Assegaff*²

Program Studi Magister Sistem Informasi, STIKOM Dinamika Bangsa, Jambi
Jl. Jendral Sudirman, Kec. Thehok, (0741) 35096

E-mail : debyfransiska123@yahoo.co.id¹, setiawanassegaff@stikom-db.ac.id²

Abstract

Large and quality population is one of the basic capital and dominant factors for national development. The large number of population data in the Pasar Jambi Distric has not produced a useful knowledge for the sub-district. This study will classify the level of family welfare using population data in Pasar Jambi District by utilizing data mining science classification techniques. The classification in this study is limited to prototype. The method that will be used is the Naive Bayes method. Naive Bayes is a method in the field of probability and statistic for classifying a class. The resulting classification is Pre-prosperous, Prosperous 1 and Prosperous 2. The results of classification prediction based on training data found 180 families classified as Pre-prosperous, 2.454 families classified as Prosperous 1, and 773 families classified as Prosperous 2. The result of accuracy obtained are said to be close to 1, 0,98 with an error value of 0,02. This classification technique can facilitate family planning field officers (PLKB) in determining the level of family welfare.

Keywords : Data Mining, Classification, Naive Bayes, Family Welfare Level, Confusion Matrix.

Abstrak

Jumlah penduduk yang besar dan berkualitas merupakan salah satu modal dasar dan faktor dominan bagi pembangunan nasional. Banyaknya data penduduk pada Kecamatan Pasar Jambi belum menghasilkan suatu pengetahuan yang bermanfaat untuk kecamatan tersebut. Penelitian ini akan melakukan klasifikasi tingkat kesejahteraan keluarga menggunakan data penduduk pada Kecamatan Pasar Jambi dengan memanfaatkan ilmu data *mining* teknik klasifikasi. Klasifikasi pada penelitian ini sebatas *Prototype*. Metode yang akan digunakan adalah metode *Naive Bayes*. *Naive Bayes* merupakan sebuah metode dalam bidang probabilitas dan *statistik* untuk mengelompokkan suatu kelas. Klasifikasi yang dihasilkan yaitu Prasejahtera, Sejahtera 1 dan Sejahtera 2. Hasil prediksi klasifikasi berdasarkan data latih (*training*) didapatkan 180 keluarga tergolong Prasejahtera, 2.454 keluarga tergolong Sejahtera 1 dan 773 keluarga tergolong Sejahtera 2. Hasil akurasi yang didapatkan dikatakan baik karena mendekati 1 yaitu 0,98, dengan nilai *error* sebesar 0,02. Teknik klasifikasi ini dapat mempermudah petugas lapangan keluarga berencana (PLKB) dalam menentukan tingkat kesejahteraan keluarga.

Kata Kunci : Data Mining, Klasifikasi, Naive Bayes, Tingkat Kesejahteraan Keluarga, Matriks Konfusi

© 2019 Jurnal Manajemen Sistem Informasi.

1. PENDAHULUAN

Pendataan penduduk untuk dikelompokkan kedalam tingkat kesejahteraan keluarga dilakukan oleh PLKB (Petugas Lapangan Keluarga Berencana) di Kecamatan yang bertugas melakukan *survey* data perkepala keluarga. Data yang didapat tersebut selanjutnya digunakan sebagai acuan untuk menentukan tingkat kesejahteraan keluarga. Sering terjadi kesalahan pengelompokan suatu keluarga kedalam tingkat kesejahteraan, oleh karenanya banyak pihak yang merasa tidak puas dengan hasil yang diperoleh tersebut

dan hal tersebut berdampak pada pengambilan kebijakan yang tidak tepat sasaran. Mengingat pentingnya kualitas data dalam perencanaan dan pengembangan program keluarga sejahtera maka BKKBN perlu mengembangkan sistem pendataan yang lebih baik.

Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai *database* besar menurut Turban (Kusrini, 2009 ; 3). Dalam *data mining* digunakan banyak algoritma, antara lain algoritma C4.5 atau Pohon Keputusan, algoritma *Nearest Neighbor*, algoritma *Naive Bayes*, *Fuzzy C Means* dan masih banyak lagi algoritma lainnya.

Data mining dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan, yaitu Deskripsi, Estimasi, Prediksi, Klasifikasi, Pengklusteran dan Asosiasi. Klasifikasi dapat menggolongkan data dan memprediksi kategori label kelas berdasarkan atribut-atribut yang berpengaruh dengan algoritma tertentu. Algoritma *Naive Bayes* merupakan sebuah metode dalam bidang probabilitas dan statistik yang sangat efektif dan efisien dan banyak digunakan oleh peneliti lainnya. Berdasarkan latar belakang tersebut penulis tertarik melakukan penelitian yang berjudul “**PENENTUAN KLASIFIKASI TINGKAT KESEJAHTERAAN KELUARGA MENGGUNAKAN METODE NAIVE BAYES PADA KECAMATAN PASAR JAMBI**”.

2. TINJAUAN PUSTAKA/PENELITIAN SEBELUMNYA

2.1 Keluarga Sejahtera

Menurut Undang-Undang No.10 (1992 ; 3) : “Keluarga Sejahtera adalah keluarga yang dibentuk berdasarkan atas perkawinan yang sah, mampu memenuhi kebutuhan hidup spiritual dan materiil yang layak, bertaqwa kepada Tuhan Yang Maha Esa, memiliki hubungan yang serasi, selaras, dan seimbang antar anggota dan antar keluarga dengan masyarakat dan lingkungan”.

2.2 Tingkat Kesejahteraan Keluarga

Tingkat kesejahteraan keluarga dikelompokkan menjadi 5 (lima) tahapan (BKKBN, 2006) yaitu :

1. Keluarga Prasejahtera (KPS), yaitu keluarga-keluarga yang belum memenuhi kebutuhan dasarnya (*basic needs*) secara minimal seperti kebutuhan akan pangan, sandang, papan, kesehatan dan pendidikan.
2. Keluarga Sejahtera I (KS I), yaitu keluarga-keluarga yang telah memenuhi kebutuhan dasarnya secara minimal, tetapi belum dapat memenuhi seluruh kebutuhan sosial psikologisnya (*socio psychological needs*), seperti kebutuhan ibadah, makan protein hewani, pakaian, ruang untuk interaksi keluarga dalam keadaan sehat, mempunyai penghasilan, bisa baca tulis latih dan keluarga berencana.
3. Keluarga Sejahtera II (KS II), yaitu keluarga-keluarga yang disamping telah dapat memenuhi kebutuhan dasarnya juga telah dapat memenuhi seluruh kebutuhan psikologinya, akan tetapi belum dapat memenuhi keseluruhan kebutuhan perkembangannya (*developmental needs*) seperti kebutuhan untuk peningkatan agama, menabung, berinteraksi dalam keluarga, ikut melaksanakan kegiatan masyarakat dan memperoleh informasi.
4. Keluarga Sejahtera III (KS III), yaitu keluarga yang telah dapat memenuhi seluruh kebutuhan dasar, kebutuhan sosial psikologisnya dan kebutuhan perkembangannya namun belum dapat memberikan sumbangan (kontribusi) yang maksimal terhadap masyarakat, seperti secara teratur (waktu tertentu) memberikan sumbangan dalam bentuk material dan keuangan untuk kepentingan sosial masyarakat atau berperan serta aktif dengan menjadi pengurus lembaga kemasyarakatan atau yayasan sosial, keagamaan, kesenian, olahraga, pendidikan dan sebagainya.
5. Keluarga Sejahtera III Plus (KS III Plus) yaitu keluarga-keluarga yang telah dapat memenuhi seluruh kebutuhannya, baik yang bersifat dasar, sosial psikologis maupun yang bersifat pengembangan serta telah dapat pula memberikan sumbangan yang nyata dan berkelanjutan bagi masyarakat.

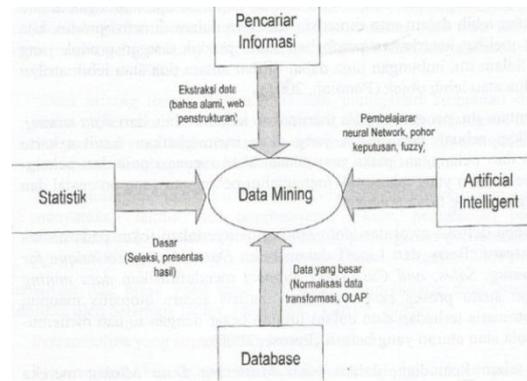
2.3 Data

Menurut Gordon B.Davis (Hutahaean, 2015 ; 8), Data adalah bahan mentah bagi informasi, dirumuskan sebagai kelompok lambang-lambang tidak acak menunjukkan jumlah-jumlah, tindakan-tindakan, hal-hal dan sebagainya. Selanjutnya Hutahaean menyebutkan metode pengumpulan data dapat

melalui pengamatan sendiri secara langsung, wawancara, perkiraan korespondensi serta melalui daftar pertanyaan.

2.4 Data Mining

Menurut Turban (Kusrini dan Luthfi, 2009 ; 3), *Data Mining* adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai *database* besar.



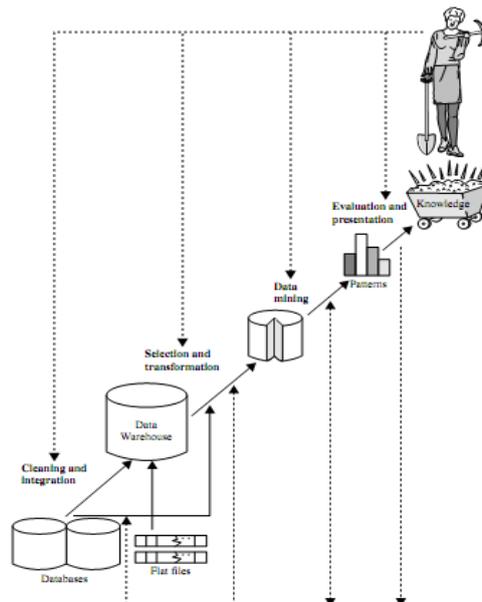
Gambar 2.1 Bidang Ilmu Data Mining (Kusrini dan Luthfi, 2009 ; 6)

Gambar diatas menunjukkan *data mining* memiliki akar yang panjang dari bidang ilmu seperti kecerdasan buatan (*artificial intelligence*), *machine learning*, statistik, *database*, dan juga *information retrieval*.

Istilah *data mining* dan *knowledge discovery in database* (KDD) sering kali digunakan secara bergantian untuk menjelaskan proses penggalian informasi tersembunyi dalam suatu basis data yang besar. Sebenarnya kedua istilah tersebut memiliki konsep yang berbeda, tetapi berkaitan satu sama lain. Dan salah satu tahapan dalam keseluruhan proses KDD adalah *data mining*.

2.5 Tahap-Tahap Data Mining

Langkah penting dalam proses penemuan pengetahuan digambarkan seperti pada Gambar 2.2 dibawah ini.



Gambar 2.2 Tahap-Tahap Data Mining (Han, dkk, 2012 ;7)

1. Pembersihan Data (*Data Cleaning*)

Pada umumnya data yang diperoleh, baik dari database suatu perusahaan maupun hasil eksperimen, memiliki isian-isian yang tidak sempurna seperti data yang hilang, data yang tidak valid atau juga hanya sekedar salah ketik. Selain itu, ada juga atribut-atribut data yang tidak relevan dengan hipotesa data *mining* yang kita miliki. Pembersihan data juga akan mempengaruhi performansi dari sistem *data mining* karena data yang ditangani akan berkurang jumlah dan kompleksitasnya.

2. Integrasi Data (*Data Integration*)

Integrasi data dilakukan pada atribut-atribut yang mengidentifikasi entitas-entitas yang unik seperti atribut nama, jenis produk, nomor pelanggan. Integrasi data perlu dilakukan secara cermat karena kesalahan pada integrasi data bisa menghasilkan hasil yang menyimpang dan bahkan menyesatkan pengambilan aksi nantinya. Sebagai contoh bila integrasi data berdasarkan jenis produk ternyata menggabungkan produk dari kategori yang berbeda maka akan didapatkan korelasi antar produk yang sebenarnya tidak ada.

3. Pemilihan Data (*Data Selection*)

Data yang ada pada *database* seringkali tidak semuanya dipakai, oleh karena itu hanya data yang sesuai untuk dianalisis yang akan diambil dari *database*. Sebagai contoh, sebuah kasus yang meneliti faktor kecenderungan orang membeli dalam kasus *market basket analysis*, tidak perlu mengambil nama pelanggan, cukup dengan id pelanggan saja.

4. Transformasi Data (*Data Transformation*)

Beberapa teknik *data mining* membutuhkan format data yang khusus sebelum bisa diaplikasikan. Sebagai contoh beberapa teknik standar seperti analisis asosiasi dan *clustering* hanya bisa menerima input data kategorikal. Karenanya data berupa angka numerik yang berlanjut perlu dibagi-bagi menjadi beberapa interval. Proses ini sering disebut *binning*.

5. Penambangan Data (*Data Mining*)

Data mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu.

6. Evaluasi Pola (*Pattern Evaluation*)

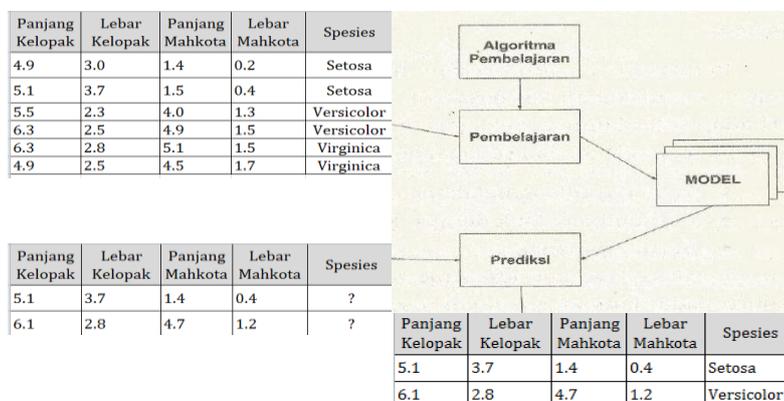
Untuk mengidentifikasi pola-pola menarik ke dalam *knowledge based* yang ditemukan. Dalam tahap ini hasil dari teknik *data mining* berupa pola-pola yang khas maupun model prediksi dievaluasi untuk menilai apakah hipotesa yang ada memang tercapai. Bila ternyata hasil yang diperoleh tidak sesuai hipotesa ada beberapa alternatif yang dapat diambil seperti menjadikannya umpan balik untuk memperbaiki proses *data mining*, mencoba metode *data mining* lain yang lebih sesuai, atau menerima hasil ini sebagai suatu hasil yang di luar dugaan yang mungkin bermanfaat.

7. Presentasi Pengetahuan (*Knowledge Presentation*)

Tahap terakhir dari proses *data mining* adalah bagaimana memformulasikan keputusan atau aksi dari hasil analisa yang didapat. Ada kalanya hal ini harus melibatkan orang-orang yang tidak memahami *data mining*. Karenanya presentasi hasil *data mining* dalam bentuk pengetahuan yang bisa dipahami semua orang adalah satu tahapan yang diperlukan dalam proses *data mining*.

2.6 Klasifikasi

Menurut Zaki et al (Suyanto, 2017 ; 115), Klasifikasi dapat didefinisikan sebagai proses untuk menyatakan suatu objek data sebagai salah satu kategori (kelas) yang telah didefinisikan sebelumnya.



Gambar 2.3 Skema Dasar Proses Klasifikasi (Kusrini dan Luthfi, 2009 ; 55)

Gambar 2.3 diatas menjelaskan proses dasar dari permasalahan klasifikasi. Sekumpulan data yang telah diketahui kelasnya digunakan sebagai data latih (*training set*). Data latih digunakan oleh suatu algoritma klasifikasi untuk proses pembelajaran akan mengenali relasi antar kelas dan variabel-variabel pada data latih. Pembelajaran yang menggunakan contoh data yang telah diketahui kelasnya disebut dengan *supervised learning*. Relasi yang telah dikenali disebut dengan model. Model dapat dimanfaatkan pada proses prediksi untuk menentukan kelas dari data uji (*test set*), yaitu sekumpulan data baru yang belum diketahui kelasnya.

2.7 Naive Bayes

Naive Bayes atau metode *Bayesian Classifier* menggunakan pendekatan teori peluang untuk melakukan klasifikasi. Metode yang menjadi acuan pada *Bayesian Classifier* adalah Teorema Bayes. Teorema Bayes menjelaskan peluang suatu kejadian berdasarkan kombinasi antara pengetahuan sebelumnya tentang kejadian tersebut dan bukti-bukti baru yang dikumpulkan dari data. Secara matematis, Teorema Bayes menurut dapat diekspresikan sebagai berikut (Adinugroho dan Sari, 2018 ; 73) :

$$\frac{P(A|B) = P(A)P(B|A)}{P(B)} \dots\dots\dots(2.1)$$

Dalam hal ini :

- P(A|B) : Peluang terjadinya kejadian A dengan syarat kejadian B telah terjadi
- P(B|A) : Peluang terjadinya kejadian B dengan syarat kejadian A telah terjadi
- P(A) : Peluang terjadinya kejadian A, tanpa pengaruh kejadian yang lain
- P(B) : Peluang terjadinya kejadian B, tanpa pengaruh kejadian yang lain

Persamaan Teorema Bayes tersebut dapat dimanfaatkan untuk melakukan klasifikasi. Pada permasalahan klasifikasi, variabel B pada persamaan Teorema Bayes menyatakan sekumpulan variabel-variabel pada data sedangkan A menyatakan kelas pada data. Nilai P(A|B) biasa disebut peluang posterior dan P(A) disebut sebagai peluang prior.

Selama proses *training*, probabilitas posterior P(A|B) ditelaah dari setiap kombinasi A dan B dari data latih. Berdasarkan nilai probabilitas posterior, sebuah data uji B' dapat diklasifikasikan ke dalam kelas A' yang memiliki nilai probabilitas posterior P(A'|B') tertinggi.

2.8 Evaluasi Metode Klasifikasi

Sebuah sistem yang melakukan klasifikasi diharapkan dapat melakukan klasifikasi dengan benar, tetapi tidak dipungkiri bahwa kinerja suatu sistem tidak bisa 100% benar sehingga sebuah sistem klasifikasi juga harus diukur kinerjanya. Umumnya, pengukuran kinerja klasifikasi dilakukan dengan matriks konfusi (*confusion matrix*). Matriks konfusi merupakan tabel pencatat hasil kerja klasifikasi.

Tabel 2.1 Tabel Confusion Matrix (Adinugroho dan Sari, 2018 ; 88)

		Nilai yang sebenarnya	
		TRUE	FALSE
Nilai Prediksi	TRUE	TP <i>(True Positive)</i>	FP <i>(False Positive)</i>
	FALSE	FN <i>(False Negative)</i>	TN <i>(True Negative)</i>

Keterangan Tabel 2.1 :

- TP : hasil prediksi sesuai dengan nilai sebenarnya yaitu sama-sama benar
- TN : hasil prediksi salah dan nilai yang sebenarnya juga salah
- FP : hasil prediksi benar, namun hasil yang sebenarnya salah
- FN : hasil prediksi salah, namun hasil yang sebenarnya benar

Beberapa metode evaluasi dapat diturunkan dari terminologi *confusion matrix*, menurut Adinugroho dan Sari (2018 ; 89-90) sebagai berikut :

1. Akurasi, merupakan hasil perhitungan semua nilai prediksi yang benar dibagi dengan keseluruhan data. Nilai akurasi terbaik jika nilai akurasi tersebut sama dengan 1.0 dan nilai paling buruk adalah 0.0.

$$\text{Akurasi} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FN} + \text{FP}} \dots\dots\dots (2.2)$$

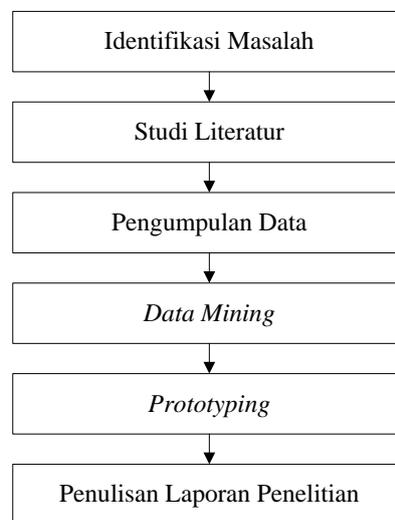
2. *Error Rate* (ER), adalah hasil perhitungan semua nilai yang dengan prediksi salah dibagi dengan keseluruhan data. Nilai akurasi terbaik merupakan kebalikan dari akurasi, yaitu 0, karena nilai yang dibandingkan adalah nilai yang prediksinya salah. Sementara nilai akurasi 1 untuk yang paling buruk.

$$\text{ER} = \frac{\text{FP} + \text{FN}}{\text{TP} + \text{TN} + \text{FN} + \text{FP}} \dots\dots\dots (2.3)$$

3. METODOLOGI

3.1 Alur Penelitian

Dalam menyelesaikan penelitian ini dilakukan beberapa tahapan yang peneliti lakukan melalui kerangka kerja penelitian, yaitu :



Gambar 3.1 Kerangka Kerja Penelitian

3.2 Bahan Penelitian

Bahan penelitian yang dibutuhkan dalam Mengklasifikasikan Tingkat Kesejahteraan Keluarga Pada Kecamatan Pasar Jambi adalah sebagai berikut :

1. Profil Balai KB di Kecamatan Pasar Jambi
2. Data Keluarga pada Kecamatan Pasar Jambi
3. Masalah dan Kendala dalam sistem yang sedang berjalan
4. Undang Undang RI Nomor 10 Tahun 1992
5. Infrastruktur Teknologi Informasi

4. HASIL DAN PEMBAHASAN

4.1 Pengujian Metode Klasifikasi Menggunakan Data *Training*

Jumlah data yang peneliti dapatkan berjumlah 3.567 data. Kemudian data tersebut dibagi 2 (dua) menjadi data latih (*training*) dan data uji (*testing*). Data latih diambil dari 3 kelurahan berjumlah 3.407 data, sementara data uji diambil dari 1 kelurahan berjumlah 160 data.

Berikut hasil prediksi klasifikasi dari data *latih* menggunakan *tools weka*.

Relation: terbaru_predicted																							
No	1: C1	2: C2	3: C3	4: C4	5: C5	6: C6	7: C7	8: C8	9: C9	10: C10	11: C11	12: C12	13: C13	14: C14	15: C15	16: C16	17: C17	18: C18	19: C19	20: C20	21: prediction margin	22: predicted Tingkat KS	23: Tingkat KS
	Nominal	Nominal	Nominal																				
1	SLTP	Wir...	Ren...	KAW...	BPJ...	Me...	Me...	Me...	Me...	Mem...	Mem...	Peng...	Mem...	Mem...	Mem...	Mem...	Mem...	seng...	Besar	Sedikit	0.916478	Sejahtera 1	Sejahtera 1
2	SLTA	Ped...	Sed...	KAW...	Tida...	Me...	Me...	Me...	Me...	Mem...	Mem...	Peng...	Mem...	Mem...	Mem...	Mem...	Mem...	seng...	Besar	Seda...	0.995095	Sejahtera 1	Sejahtera 1
3	SLTP	Ped...	Sed...	KAW...	Tida...	Me...	Me...	Me...	Me...	Mem...	Mem...	Peng...	Mem...	Mem...	Mem...	Mem...	Tida...	seng...	Besar	Sedikit	0.99729	Sejahtera 1	Sejahtera 1
4	SLTA	Wir...	Sed...	KAW...	BPJ...	Me...	Me...	Me...	Me...	Mem...	Mem...	Peng...	Mem...	Mem...	Mem...	Mem...	Tida...	seng...	Besar	Sedikit	0.995815	Sejahtera 1	Sejahtera 1
5	Aka...	Lain...	Sed...	KAW...	Tida...	Me...	Me...	Me...	Me...	Mem...	Mem...	Tida...	Mem...	Mem...	Mem...	Mem...	Mem...	seng...	Besar	Seda...	0.998827	Sejahtera 1	Sejahtera 1
6	Aka...	Lain...	Sed...	KAW...	Tida...	Me...	Me...	Me...	Me...	Mem...	Mem...	Peng...	Mem...	Mem...	Mem...	Mem...	Mem...	seng...	Besar	Sedikit	0.996593	Sejahtera 1	Sejahtera 1
7	SLTA	Wir...	Sed...	KAW...	BPJ...	Me...	Me...	Me...	Me...	Mem...	Mem...	Tida...	Mem...	Mem...	Mem...	Mem...	Mem...	seng...	Besar	Seda...	0.998411	Sejahtera 1	Sejahtera 1
8	SLTA	Wir...	Sed...	KAW...	Tida...	Me...	Me...	Me...	Me...	Mem...	Mem...	Tida...	Mem...	Mem...	Mem...	Mem...	Mem...	seng...	Besar	Seda...	0.99887	Sejahtera 1	Sejahtera 1
9	SLTA	Peg...	Tinggi	KAW...	Tida...	Me...	Me...	Me...	Me...	Mem...	seng...	Besar	Seda...	0.999099	Sejahtera 2	Sejahtera 2							
10	SLTA	Wir...	Sed...	KAW...	Tida...	Me...	Me...	Me...	Me...	Mem...	Mem...	Peng...	Mem...	Mem...	Mem...	Mem...	Tida...	seng...	Besar	Sedikit	0.996995	Sejahtera 1	Sejahtera 1
11	SD	Pek...	Sed...	KAW...	BPJ...	Me...	Me...	Me...	Me...	Mem...	seng...	Besar	Seda...	0.539152	Sejahtera 1	Sejahtera 1							

Dari hasil klasifikasi yang disajikan diatas, dapat diperoleh tabel *confusion matrix* sebagai berikut :

Tabel 4.1 Tabel *Confusion Matrix* Hasil Klasifikasi Data Training

3407 record	Sejahtera 1	Sejahtera 2	Prasejahtera	Jumlah
Sejahtera 1	2403	3	1	2407
Sejahtera 2	49	769	14	832
Prasejahtera	2	1	165	168
Jumlah	2454	773	180	3407

Keterangan Tabel :

- Dari 3.407 data hasil klasifikasi PLKB menyatakan jumlah kategori Prasejahtera sebanyak 180 keluarga, Sejahtera 1 sebanyak 2.454 keluarga, dan Sejahtera 2 sebanyak 773 keluarga.
- Sementara, hasil prediksi klasifikasi *Naive bayes* menggunakan *tools weka* didapatkan jumlah kategori Prasejahtera sebanyak 168 keluarga, Sejahtera 1 sebanyak 2.407 keluarga, dan Sejahtera 2 sebanyak 832 keluarga.
- Prediksi berdasarkan tabel matrix konfusi diatas, dapat dilihat data yang diagonal merupakan data yang terklasifikasi benar sebanyak 3.337 sedangkan data yang belum tepat terklasifikasi sebanyak 70.
- Penjabaran prediksi data yang belum tepat klasifikasinya :
 - Kelas Sejahtera 1 yang masuk ke kelas Sejahtera 2 sebanyak 3
 - Kelas Sejahtera 1 yang masuk ke kelas Prasejahtera sebanyak 1
 - Kelas Sejahtera 2 yang masuk ke kelas Sejahtera 1 sebanyak 49
 - Kelas Sejahtera 2 yang masuk ke kelas Prasejahtera sebanyak 14
 - Kelas Prasejahtera yang masuk ke kelas Sejahtera 1 sebanyak 2
 - Kelas Prasejahtera yang masuk ke kelas Sejahtera 2 sebanyak 1
- Nilai Akurasi = $(TP+TN) / \text{TOTAL DATA} = (2.403 + 769 + 165) / 3.407 = 3.337 / 3.407 = 0,98$. Ini berarti nilai akurasi dari klasifikasi cukup tinggi karena mendekati 1.
- Nilai *Error Rate* (ER) = $FP+FN / \text{TOTAL DATA} = (3 + 1 + 49 + 14 + 2 + 1) / 3.407 = 70 / 3.407 = 0,02$. Ini berarti nilai eror dari hasil klasifikasi cukup rendah karena mendekati 0.

4.2 Perhitungan Manual Metode Klasifikasi

Untuk pengujian manual peneliti mengambil sampel data dari data latih, sementara data uji digunakan peneliti untuk memprediksi data baru yang belum terklasifikasi. Berikut langkah perhitungan manual dengan rumus *Naive bayes* menggunakan sampel data baru yang belum terklasifikasi dengan kriteria seperti tabel dibawah ini.

Tabel 4.2 Data Baru Untuk Di Testing

Kode	Kriteria	Prasejahtera	Sejahtera 1	Sejahtera 2	Prasejahtera	Sejahtera 1	Sejahtera 2
C1	SLTP	27	508	171	0,15	0,21	0,22
C2	Pedagang	38	385	126	0,21	0,16	0,16
C3	Sedang	1	2411	34	0,01	0,98	0,04
C4	Belum Kawin	10	159	50	0,06	0,06	0,06
C5	BPJS-PBI	41	397	205	0,23	0,16	0,27
C6	Memenuhi	152	2454	773	0,84	1,00	1,00
C7	Memenuhi	132	2454	773	0,73	1,00	1,00
C8	Memenuhi	152	2454	773	0,84	1,00	1,00
C9	Memenuhi	152	2454	773	0,84	1,00	1,00
C10	Memenuhi	152	2454	773	0,84	1,00	1,00
C11	Memenuhi	152	2454	773	0,84	1,00	1,00
C12	Pengecualian	66	2182	73	0,37	0,89	0,09
C13	Tidak Memenuhi	69	58	12	0,38	0,02	0,02
C14	Memenuhi	111	2442	773	0,62	1,00	1,00
C15	Tidak Memenuhi	28	234	50	0,16	0,10	0,06
C16	Tidak Memenuhi	68	13	14	0,38	0,01	0,02
C17	Memenuhi	49	1519	507	0,27	0,62	0,66
C18	Genteng/tembok/semen	33	349	138	0,18	0,14	0,18
C19	Kecil	11	334	145	0,06	0,14	0,19
C20	Sedang	85	795	339	0,47	0,32	0,44

Berikut proses perhitungan manual menggunakan formula *Naive bayes* :

- a. Menghitung jumlah *class* / label
 $P(Y=\text{Prasejahtera}) = 168/3407 = 0,05$
 “Jumlah data Prasejahtera pada data latih dibagi dengan jumlah keseluruhan data latih”
 $P(Y=\text{Sejahtera 1}) = 2407/3407 = 0,71$
 “Jumlah data Sejahtera 1 pada data latih dibagi dengan jumlah keseluruhan data latih”
 $P(Y=\text{Sejahtera 2}) = 832/3407 = 0,24$
 “Jumlah data Sejahtera 2 pada data latih dibagi dengan jumlah keseluruhan data latih”
- b. Menghitung jumlah kasus yang sama dengan *class* yang sama
 $P(C1=\text{SLTP} | Y=\text{Prasejahtera}) = 27/168 = 0,15$
 $P(C1=\text{SLTP} | Y=\text{Sejahtera 1}) = 508/2407 = 0,21$
 $P(C1=\text{SLTP} | Y=\text{Sejahtera 2}) = 198/832 = 0,22$
 $P(C2=\text{Pedagang} | Y=\text{Prasejahtera}) = 38/168 = 0,21$
 $P(C2=\text{Pedagang} | Y=\text{Sejahtera 1}) = 385/2407 = 0,16$
 $P(C2=\text{Pedagang} | Y=\text{Sejahtera 2}) = 126/832 = 0,16$
 $P(C3=\text{Sedang} | Y=\text{Prasejahtera}) = 1/168 = 0,01$
 $P(C3=\text{Sedang} | Y=\text{Sejahtera 1}) = 2411/2407 = 0,98$
 $P(C3=\text{Sedang} | Y=\text{Sejahtera 2}) = 34/832 = 0,04$

$P(C4=Belum Kawin | Y=Prasejahtera) = 10/168 = 0,06$
 $P(C4=Belum Kawin | Y=Sejahtera 1) = 159/2407 = 0,06$
 $P(C4=Belum Kawin | Y= Sejahtera 2) = 50/832 = 0,06$
 $P(C5=BPJS-PBI | Y=Prasejahtera) = 41/168 = 0,23$
 $P(C5=BPJS-PBI | Y=Sejahtera 1) = 397/2407 = 0,16$
 $P(C5=BPJS-PBI | Y= Sejahtera 2) = 205/832 = 0,27$
 $P(C6=Memenuhi | Y=Prasejahtera) = 152/168 = 0,84$
 $P(C6=Memenuhi | Y=Sejahtera 1) = 2454/2407 = 1,00$
 $P(C6=Memenuhi | Y= Sejahtera 2) = 773/832 = 1,00$
 $P(C7=Memenuhi | Y=Prasejahtera) = 132/168 = 0,73$
 $P(C7=Memenuhi | Y=Sejahtera 1) = 2454/2407 = 1,00$
 $P(C7=Memenuhi | Y= Sejahtera 2) = 773/832 = 1,00$
 $P(C8=Memenuhi | Y=Prasejahtera) = 152/168 = 0,84$
 $P(C8=Memenuhi | Y=Sejahtera 1) = 2454/2407 = 1,00$
 $P(C8=Memenuhi | Y= Sejahtera 2) = 773/832 = 1,00$
 $P(C9=Memenuhi | Y=Prasejahtera) = 152/168 = 0,84$
 $P(C9=Memenuhi | Y=Sejahtera 1) = 2454/2407 = 1,00$
 $P(C9=Memenuhi | Y= Sejahtera 2) = 773/832 = 1,00$
 $P(C10=Memenuhi | Y=Prasejahtera) = 152/168 = 0,84$
 $P(C10=Memenuhi | Y=Sejahtera 1) = 2454/2407 = 1,00$
 $P(C10=Memenuhi | Y= Sejahtera 2) = 773/832 = 1,00$
 $P(C11=Memenuhi | Y=Prasejahtera) = 152/168 = 0,84$
 $P(C11=Memenuhi | Y=Sejahtera 1) = 2454/2407 = 1,00$
 $P(C11=Memenuhi | Y= Sejahtera 2) = 773/832 = 1,00$
 $P(C12=Pengecualian | Y=Prasejahtera) = 66/168 = 0,37$
 $P(C12=Pengecualian | Y=Sejahtera 1) = 2182/2407 = 0,89$
 $P(C12=Pengecualian | Y= Sejahtera 2) = 73/832 = 0,09$
 $P(C13=Tidak Memenuhi | Y=Prasejahtera) = 69/168 = 0,38$
 $P(C13=Tidak Memenuhi | Y=Sejahtera 1) = 58/2407 = 0,02$
 $P(C13=Tidak Memenuhi | Y= Sejahtera 2) = 12/832 = 0,02$
 $P(C14=Memenuhi | Y=Prasejahtera) = 111/168 = 0,62$
 $P(C14=Memenuhi | Y=Sejahtera 1) = 2442/2407 = 1,00$
 $P(C14=Memenuhi | Y= Sejahtera 2) = 773/832 = 1,00$
 $P(C15=Tidak Memenuhi | Y=Prasejahtera) = 28/168 = 0,16$
 $P(C15=Tidak Memenuhi | Y=Sejahtera 1) = 234/2407 = 0,10$
 $P(C15=Tidak Memenuhi | Y= Sejahtera 2) = 50/832 = 0,06$
 $P(C16=Tidak Memenuhi | Y=Prasejahtera) = 68/168 = 0,38$
 $P(C16=Tidak Memenuhi | Y=Sejahtera 1) = 13/2407 = 0,01$
 $P(C16=Tidak Memenuhi | Y= Sejahtera 2) = 14/832 = 0,02$
 $P(C17=Memenuhi | Y=Prasejahtera) = 49/168 = 0,27$
 $P(C17=Memenuhi | Y=Sejahtera 1) = 1519/2407 = 0,62$
 $P(C17=Memenuhi | Y= Sejahtera 2) = 507/832 = 0,66$
 $P(C18=Genteng/Tembok/Semen | Y=Prasejahtera) = 33/168 = 0,18$
 $P(C18= Genteng/Tembok/Semen | Y=Sejahtera 1) = 349/2407 = 0,14$
 $P(C18= Genteng/Tembok/Semen | Y= Sejahtera 2) = 138/832 = 0,18$
 $P(C19=Kecil | Y=Prasejahtera) = 11/168 = 0,06$
 $P(C19= Kecil | Y=Sejahtera 1) = 334/2407 = 0,14$
 $P(C19= Kecil | Y= Sejahtera 2) = 145/832 = 0,19$
 $P(C20=Sedang | Y=Prasejahtera) = 85/168 = 0,47$
 $P(C20=Sedang | Y=Sejahtera 1) = 795/2407 = 0,32$
 $P(C20= Sedang | Y= Sejahtera 2) = 339/832 = 0,44$

c. Kalikan semua hasil variabel Prasejahtera, Sejahtera 1 dan Sejahtera 2

$P(C1=SLTP | Y=Prasejahtera) \times P(C2=Pedagang | Y=Prasejahtera) \times P(C3=Sedang | Y=Prasejahtera) \times P(C4=Belum Kawin | Y=Prasejahtera) \times P(C5=BPJS-PBI | Y=Prasejahtera) \times P(C6=Memenuhi | Y=Prasejahtera) \times P(C7= Memenuhi | Y=Prasejahtera) \times P(C8=Memenuhi | Y=Prasejahtera) \times P(C9=Memenuhi | Y=Prasejahtera) \times P(C10=Memenuhi | Y=Prasejahtera) \times P(C11=Memenuhi | Y=Prasejahtera) \times$

$$\begin{aligned}
 &P(C12=Pengeceualian | Y=Prasejahtera) \times P(C13=Tidak Memenuhi | Y=Prasejahtera) \times \\
 &P(C14= Memenuhi | Y=Prasejahtera) \times P(C15= Tidak Memenuhi | Y=Prasejahtera) \times \\
 &P(C16=Tidak Memenuhi | Y=Prasejahtera) \times P(C17=Memenuhi | Y=Prasejahtera) \times \\
 &P(C18=Genteng/Tembok/Semen | Y=Prasejahtera) \times P(C19=Kecil | Y=Prasejahtera) \times P \\
 &(C20=Sedang | Y=Prasejahtera) \\
 &= 0,15 \times 0,21 \times 0,01 \times 0,06 \times 0,23 \times 0,84 \times 0,73 \times 0,84 \times 0,84 \times 0,84 \times 0,37 \times 0,38 \times \\
 &0,62 \times 0,16 \times 0,38 \times 0,27 \times 0,18 \times 0,06 \times 0,47 \\
 &= 9,63995258E-12
 \end{aligned}$$

$$\begin{aligned}
 &P(C1=SLTP | Y=Sejahtera 1) \times P(C2=Pedagang | Y= Sejahtera 1) \times P(C3=Sedang | Y= \\
 &Sejahtera 1) \times P(C4=Belum Kawin | Y= Sejahtera 1) \times P(C5=BPJS-PBI | Y= Sejahtera 1) \times \\
 &P(C6=Memenuhi | Y= Sejahtera 1) \times P(C7= Memenuhi | Y= Sejahtera 1) \times P(C8=Memenuhi \\
 &| Y= Sejahtera 1) \times P(C9=Memenuhi | Y= Sejahtera 1) \times P(C10=Memenuhi | Y= Sejahtera 1) \\
 &\times P(C11=Memenuhi | Y= Sejahtera 1) \times P(C12=Pengeceualian | Y= Sejahtera 1) \times \\
 &P(C13=Tidak Memenuhi | Y= Sejahtera 1) \times P(C14= Memenuhi | Y= Sejahtera 1) \times P(C15= \\
 &Tidak Memenuhi | Y= Sejahtera 1) \times P(C16=Tidak Memenuhi | Y= Sejahtera 1) \times \\
 &P(C17=Memenuhi | Y= Sejahtera 1) \times P(C18=Genteng/Tembok/Semen | Y= Sejahtera 1) \times \\
 &P(C19=Kecil | Y= Sejahtera 1) \times P (C20=Sedang | Y= Sejahtera 1) \\
 &= 0,21 \times 0,16 \times 0,98 \times 0,06 \times 0,16 \times 1,00 \times 1,00 \times 1,00 1,00 \times 1,00 \times 1,00 \times 0,89 \times 0,02 \times \\
 &1,00 \times 0,10 \times 0,01 \times 0,62 \times 0,14 \times 0,14 \times 0,32 \\
 &= 1,36752207E-10
 \end{aligned}$$

$$\begin{aligned}
 &P(C1=SLTP | Y=Sejahtera 2) \times P(C2=Pedagang | Y= Sejahtera 2) \times P(C3=Sedang | Y= \\
 &Sejahtera 2) \times P(C4=Belum Kawin | Y= Sejahtera 2) \times P(C5=BPJS-PBI | Y= Sejahtera 2) \times \\
 &P(C6=Memenuhi | Y= Sejahtera 2) \times P(C7= Memenuhi | Y= Sejahtera 2) \times P(C8=Memenuhi \\
 &| Y= Sejahtera 2) \times P(C9=Memenuhi | Y= Sejahtera 2) \times P(C10=Memenuhi | Y= Sejahtera 2) \\
 &\times P(C11=Memenuhi | Y= Sejahtera 2) \times P(C12=Pengeceualian | Y= Sejahtera 2) \times \\
 &P(C13=Tidak Memenuhi | Y= Sejahtera 2) \times P(C14= Memenuhi | Y= Sejahtera 2) \times P(C15= \\
 &Tidak Memenuhi | Y= Sejahtera 2) \times P(C16=Tidak Memenuhi | Y= Sejahtera 2) \times \\
 &P(C17=Memenuhi | Y= Sejahtera 2) \times P(C18=Genteng/Tembok/Semen | Y= Sejahtera 2) \times \\
 &P(C19=Kecil | Y= Sejahtera 2) \times P (C20=Sedang | Y= Sejahtera 2) \\
 &= 0,22 \times 0,16 \times 0,04 \times 0,06 \times 0,27 \times 1,00 \times 1,00 \times 1,00 \times 1,00 \times 1,00 \times 0,09 \times 0,02 \times \\
 &1,00 \times 0,06 \times 0,02 \times 0,66 \times 0,18 \times 0,19 \times 0,44 \\
 &= 4,893213E-13
 \end{aligned}$$

- d. Bandingkan hasil *class* Prasejahtera, Sejahtera 1 dan Sejahtera 2 dari hasil dari hasil diatas, terlihat bahwa nilai probabilitas tertinggi ada pada kelas (P)Prasejahtera) sehingga dapat disimpulkan bahwa penduduk tersebut masuk dalam klasifikasi “Prasejahtera”.

4.3 Pengujian Data Testing

Berdasarkan model data *training* diatas selanjutnya dapat diujikan data yang belum terklasifikasi. Peneliti menggunakan 160 data baru yang digunakan sebagai data *testing* untuk diketahui klasifikasinya.

Berikut hasil prediksi klasifikasi dari data *testing* menggunakan *tools weka*.

Relation: terbaru_predicted																								
No.	1: C1	2: C2	3: C3	4: C4	5: C5	6: C6	7: C7	8: C8	9: C9	10: C10	11: C11	12: C12	13: C13	14: C14	15: C15	16: C16	17: C17	18: C18	19: C19	20: C20	21: prediction	margin	22: predicted	Tingkat KE
	Nominal	Numeric		Nominal																				
1	Aka...	Peg...		KAW...	Tida...	Me...	Me...	Me...	Me...	Mem...	Mem...	Tida...	Mem...	Mem...	Mem...	Mem...	Mem...	seng...	Besar	Seda...	0.989334		Sejahtera 1	
2	SLTA	Peg...		KAW...	Tida...	Me...	Me...	Me...	Me...	Mem...	seng...	Besar	Seda...	-0.807916		Sejahtera 2								
3	SLTA	Peg...		KAW...	Tida...	Me...	Me...	Me...	Me...	Mem...	Mem...	Peng...	Mem...	Mem...	Mem...	Mem...	Mem...	Tida...	gent...	Besar	Sedikit	0.926615		Sejahtera 1
4	SLTA	PNS...		KAW...	BPJ...	Me...	Me...	Me...	Me...	Mem...	seng...	Besar	Seda...	-0.919806		Sejahtera 2								
5	Aka...	Peg...	Sed...	KAW...	BPJ...	Me...	Me...	Me...	Me...	Mem...	Mem...	Peng...	Mem...	Mem...	Mem...	Mem...	Mem...	Tida...	gent...	Besar	Seda...	0.99549		Sejahtera 1
6	SLTA	Lain...		JAN...	Tida...	Me...	Me...	Me...	Me...	Mem...	Mem...	Peng...	Mem...	Mem...	Mem...	Mem...	Mem...	Tida...	gent...	Besar	Sedikit	0.970146		Sejahtera 1
7	SLTA	Peg...	Sed...	KAW...	Tida...	Me...	Me...	Me...	Me...	Mem...	Mem...	Peng...	Mem...	Mem...	Mem...	Mem...	Mem...	seng...	Besar	Seda...	0.994955		Sejahtera 1	
8	SLTA	Pen...	Sed...	KAW...	BPJ...	Me...	Me...	Me...	Me...	Mem...	Mem...	Tida...	Mem...	Mem...	Mem...	Mem...	Mem...	seng...	Besar	Sedikit	0.999628		Sejahtera 1	
9	Aka...	PNS...		KAW...	BPJ...	Me...	Me...	Me...	Me...	Mem...	seng...	Besar	Seda...	-0.885712		Sejahtera 2								
10	Aka...	Lain...	Sed...	BEL...	BPJ...	Me...	Me...	Me...	Me...	Mem...	Mem...	Peng...	Mem...	Mem...	Mem...	Mem...	Mem...	Tida...	gent...	Besar	Sedikit	0.997571		Sejahtera 1
11	SLTA	Peg...		KAW...	Tida...	Me...	Me...	Me...	Me...	Mem...	Mem...	Peng...	Mem...	Mem...	Mem...	Mem...	Mem...	seng...	Besar	Sedikit	0.956831		Sejahtera 1	

4.4 Rancangan User Interface

1. Rancangan Interface Halaman Login

A login form with a blue circular icon containing a white person silhouette. Below the icon are two input fields: the first is labeled 'Username' and the second is labeled 'Password'. At the bottom of the form is a blue button labeled 'LOGIN'.

Gambar 4.3 Halaman Login

2. Rancangan Interface Halaman Data Testing

The screenshot shows a web application interface for 'PENENTUAN KLASIFIKASI TINGKAT KESEJAHTERAAN KELUARGA PADA KECAMATAN PASAR JAMBI'. The header includes a logo on the left and 'Hallo Admin' with a user icon on the right. A navigation menu on the left lists options: HOME, DATA (with a dropdown arrow), DATA ADMIN, DATA KELURAHAN, DATA KRITERIA, BOBOT KRITERIA, DATA KLASIFIKASI, DATA TRAINING, DATA TESTING, and CETAK LAPORAN. The main content area is titled 'TAMBAH DATA TESTING' and features a table with columns: NO, NIK, NAMA, TGL LAHIR, PEKERJAAN, STATUS KAWIN, and AKSI. The table contains four rows of placeholder data. Below the table are pagination controls ('Showing 1 to 4 of 160 entries', 'Previous', '1', 'Next') and two buttons: 'PROSES' (green) and 'BATAL' (yellow). A copyright notice 'CopyRight @2018 BKKBN' is at the bottom.

Gambar 4.3 Halaman Data Testing

3. Rancangan Interface Halaman Hasil Testing Data Baru

A form titled 'HASIL TESTING DATA BARU' displaying the following test results:

- NIK : 9999
- NO.KK : 9999
- Nama : XXXX
- Tanggal Lahir : dd/mmm/yyyy
- Jenis Kelamin : XXXX
- Agama : XXXX
- Tingkat Kesejahteraan : XXXX

At the bottom of the form are two buttons: 'DETAIL' (blue) and 'KEMBALI' (yellow).

Gambar 4.3 Hasil Testing Data Baru

5. PENUTUP

5.1 Kesimpulan

1. Penelitian ini menghasilkan klasifikasi tingkat kesejahteraan keluarga yaitu Prasejahtera, Sejahtera 1 dan Sejahtera 2 dengan memanfaatkan Ilmu Data *Mining* teknik klasifikasi menggunakan algoritma Naive Bayes.
2. Hasil prediksi klasifikasi berdasarkan data latih (*training*) didapatkan :
 - a. Keluarga Prasejahtera sebanyak 180 keluarga
 - b. Keluarga Sejahtera 1 sebanyak 2.454 keluarga
 - c. Keluarga Sejahtera 2 sebanyak 773 keluarga
3. Hasil akurasi dikatakan baik apabila mendekati 1 atau 100%, berbanding terbalik dengan nilai *error* dikatakan baik apabila mendekati 0, dan pada penelitian ini didapatkan nilai akurasi pada sistem yaitu 97,94% atau 0,98 dengan nilai *error* sebesar 2,05% atau 0,02.
4. Teknik Klasifikasi tingkat kesejahteraan keluarga dengan memanfaatkan Data Mining menggunakan algoritma Naive Bayes dapat mempermudah PLKB dalam menentukan tingkat kesejahteraan keluarga.

5.2 Saran

Berikut ini beberapa saran yang bertujuan pengembangan *Prototype* Algoritma *Naive Bayes* untuk klasifikasi tingkat kesejahteraan keluarga berdasarkan indikator penentuan tingkat kesejahteraan keluarga yang telah ditetapkan dari BKKBN, sebagai berikut :

1. Untuk memperbanyak jumlah klasifikasi sesuai tingkat kesejahteraan yang ada, yaitu hingga Sejahtera 3 Plus.
2. Untuk dapat dikembangkan lebih lanjut dalam bentuk aplikasi agar dapat diimplementasikan sesuai kebutuhan yang ada.

6. DAFTAR PUSTAKA

- [1] Adinugroho, Sigit; & Sari, Yuita Arum. 2018. *Implementasi Data Mining Menggunakan Weka*. Malang : UB Press.
- [2] *Buku Saku Bagi Petugas Lapangan dan Kader Tentang Pendataan Keluarga Tahun 2006*. Jambi : BKKBN Provinsi Jambi.
- [3] Han, Jiawei; & Kamber, Michelin; & Pei, Jian. 2012. *Data Mining Concept And Techniques*. USA : Morgan Kaufmann.
- [4] Hutahaean, Japerson. 2015. *Konsep Sistem Informasi*. Yogyakarta : Deepublish.
- [5] Kusriani; & Luthfi, Emha Taufiq. 2009. *Algoritma Data Mining*. Yogyakarta : Penerbit ANDI.
- [6] Suyanto. 2017. *Data Mining Untuk Klasifikasi Dan Klasterisasi Data*. Bandung : Informatika.
- [7] Undang-Undang Republik Indonesia No.10 Tahun 1992 Tentang Perkembangan Kependudukan dan Pembangunan Keluarga Sejahtera.