



PERBANDINGAN ALGORITMA NAÏVE BAYES DAN C4.5 UNTUK KLASIFIKASI BANTUAN RUMAH SEHAT

Irawan Ibrahim^{a)}, *Hilmansyah Gani*^{b)}, *Rizal Lamusu*^{c)}, *Yulan Humolungo*^{d)}

^{abcd} *Program Studi Ilmu Komputer, Fakultas Sains dan Teknologi Universitas Muhammadiyah Gorontalo*

ABSTRACT

Classification is a technique for determining group membership based on existing data. The selection of the recipients of healthy housing assistance is only decided through village deliberation forums so that it requires accuracy in fulfilling the criteria and conditions for the recipients of healthy housing assistance. In data collection, the data obtained had 11 criteria including the name of the head of the family, occupation, age of marriage, marital status, number of family members, land ownership status, roof, ceiling, floor, wall and latrine. The testing phase uses 114 data records. The accuracy of the test results using the Naïve Bayes Algorithm and C4.5 in terms of two parameters, namely X-Validation and the amount of training data. The accuracy value of the classification results in terms of the C4.5 algorithm parameter gets the highest accuracy value with an accuracy rate of 96.51% compared to the Naïve Bayes Algorithm with an accuracy rate of 95.61%.

Keywords: Data Mining, Naïve Bayes, C4.5, Classification, Healthy Home Assistance.

Klasifikasi adalah sebuah teknik untuk menentukan keanggotaan kelompok berdasarkan data-data yang sudah ada. Penentuan pemilihan penerima bantuan rumah sehat hanya di putuskan melalui forum musyawarah desa sehingga membutuhkan ketelitian dalam pemenuhan kriteria dan syarat penerima bantuan rumah sehat. Pada pengumpulan data, data yang didapatkan memiliki 11 kriteria di antaranya yaitu nama kepala keluarga, pekerjaan, usia pernikahan status perkawinan, jumlah anggota keluarga, status kepemilikan tanah, atap, langit-langit, lantai, dinding dan jamban. Tahap pengujian menggunakan 114 record data. Keakurasian dari hasil uji coba menggunakan Algoritma Naïve Bayes dan C4.5 ditinjau dari dua parameter yaitu X-Validation dan jumlah data training. Nilai keakurasian hasil klasifikasi yang ditinjau dari parameter Algoritma C4.5 mendapatkan nilai akurasi yang tertinggi dengan tingkat akurasi 96.51% dibandingkan dengan Algoritma Naïve Bayes dengan tingkat akurasi sebesar 95,61%.

Kata Kunci : Data Mining, Naïve Bayes, C4.5, Klasifikasi, Bantuan Rumah Sehat.

1. PENDAHULUAN

Dalam perkembangan teknologi masa kini, banyaknya data menjadi sebuah permasalahan sekaligus kesempatan bagi sebuah instansi. Data bisa jadi permasalahan jika tidak disimpan, dikelola, maupun di proses dengan baik. Data yang selalu bermunculan setiap waktu akan terus menumpuk bila tidak di dokumentasikan dengan baik, sehingga data tersebut menjadi tidak berguna untuk perusahaan. Sedangkan data menjadi sebuah kesempatan apabila dapat disimpan, dikelola dan di proses menjadi lebih berarti untuk instansi tersebut.

Pengelolaan data yang sangat besarahkan melibatkan proses data mining. Data mining adalah proses maupun tahapan dalam menemukan sebuah struktur data. Struktur data tersebut dapat mengambil banyak bentuk, termasuk aturan, grafik atau jaringan, pohon maupun persamaan. Dengan menggunakan data mining maka sebuah kasus dapat dilihat trend, struktur maupun prediksinya di masa mendatang. Data mining sendiri memiliki

banyak tahapan dan teknik yang dapat di implementasikan dalam kehidupan nyata.

Dalam kasus di dunia nyata, banyak teknik dalam data mining yang dapat digunakan, salah satunya adalah teknik klasifikasi. Klasifikasi sendiri merupakan bentuk dasar dari analisis data. Klasifikasi adalah sebuah teknik untuk menentukan keanggotaan kelompok berdasarkan data-data yang sudah ada. Konsep dasar dari klasifikasi adalah beberapa data yang memiliki struktur data yang mirip akan memiliki klasifikasi yang mirip pula. Sebagai contoh penerapan klasifikasi adalah untuk membagi sebuah data pada perkreditan, dengan memilih beberapa variable yang tepat, lalu menentukan dan mengklasifikasi apakah data seorang nasabah itu nantinya akan menjadi nasabah yang membayar tepat waktu, membayar terlambat atau bahkan tidak membayar. Klasifikasi ini dapat diterapkan dalam berbagai kasus untuk membentuk sebuah aturan (*rule*) terhadap sebuah data. (Kurniawan, 2017)

Pembangunan Desa adalah upaya peningkatan kualitas hidup dan kehidupan untuk sebesar-besarnya kesejahteraan masyarakat Desa. Perencanaan pembangunan desa adalah proses tahapan kegiatan yang diselenggarakan oleh pemerintah Desa dan unsur masyarakat secara partisipatif guna pemanfaatan dan pengalokasian sumber daya desa dalam rangka mencapai tujuan pembangunan desa.

Permasalahan yang dihadapi dalam program bantuan rumah sehat dalam proses verifikasi calon penerima bantuan mengakibatkan ketidakadilan bagi calon penerima bantuan. Penentuan pemilihan penerima bantuan rumah sehat hanya diputuskan melalui forum musyawarah desa sehingga membutuhkan ketelitian dalam pemenuhan kriteria dan syarat penerimaan bantuan rumah sehat.

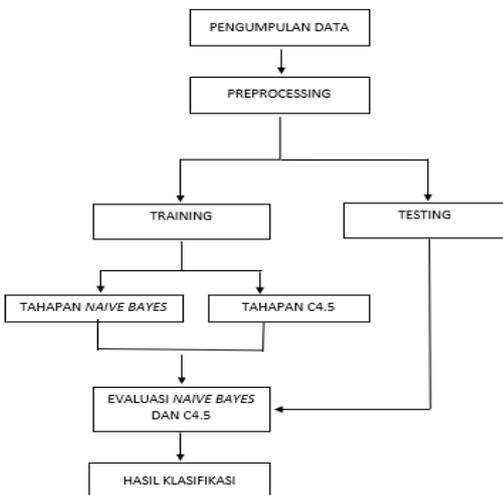
Dalam beberapa penelitian sebelumnya, telah banyak metode klasifikasi yang di implementasikan dalam kehidupan nyata. Beberapa algoritma yang sangat populer saat ini adalah *Naive Bayes* dan *C.45*. *Naive Bayes* merupakan algoritma klasifikasi dengan rumus yang sederhana dan mudah untuk di aplikasikan, sedangkan algoritma *C.45* dalam beberapa penelitian yang menggunakan *decision tree classification*, memberikan tingkat akurasi yang tinggi.

Berdasarkan beberapa penelitian-penelitian tersebut, maka penelitian ini untuk membandingkan Algoritma *Naive Bayes* serta *C4.5* untuk Bantuan Rumah Sehat dengan melakukan perhitungan *Precision*, *Recall* dan *Accuracy* untuk setiap kasus dengan perbandingan data training dan data testing yang ada. Hasil dari perhitungan *Precision*, *Recall* dan *Accuracy* di perbandingkan, sehingga dapat di ambil kesimpulan mengenai algoritma terbaik untuk klasifikasi pada permasalahan tersebut.

2. METODOLOGI

Adapun tahapan dalam penelitian ini adalah:

- a. Pengumpulan data, data yang di peroleh dari Desa Dambalo masih berupa dokumen kemudian dimasukkan ke *file excel*
- b. *Preprocessing*, data yang dilakukan pada penelitian ini adalah dengan melakukan normalisasi data dengan tujuan untuk mengelompokkan data kedalam jangkauan tertentu agar dapat memudahkan proses pengolahan data.
- c. *Training, Testing, Tahapan Naive Bayes dan Tahapan C4.5*. Data training di uji dengan menggunakan validasi untuk mendapatkan nilai *confusion matrix C4.5* dan *naive bayes* untuk mendapatkan akurasi yang terbaik.
- d. *Evaluasi C4.5 dan Naive Bayes*, Hasil dari perbandingan kedua algoritma tersebut diambil *confusion matrix*.
- e. *Hasil Klasifikasi*, diterapkan pada penentuan bantuan rumah sehat.



3. HASIL DAN PEMBAHASAN

a. Pengumpulan Data

Pada penelitian ini data yang digunakan adalah data sekunder dan data primer. Data sekunder sendiri merupakan sumber data berupa file atau berkas sedangkan data primer berupa data yang didapatkan melalui tahap wawancara. Data yang diperoleh dari Desa Dambalo, Kecamatan Tomilito yaitu sebanyak 114 data calon penerima bantuan rumah sehat.

Data asli yang di dapatkan penulis memiliki 11 kriteria diantaranya yaitu nama kepala keluarga, pekerjaan, usia pernikahan, status perkawinan, jumlah anggota keluarga, status kepemilikan tanah, atap, langit-langit, lantai, dinding dan jamban.

b. Data seleksi

Data seleksi adalah proses menganalisis data-data yang relevan dari *data base* karena tidak semua data dibutuhkan dalam proses *data mining*. Data tersebut dipilih dan diseleksi dari *database* untuk dianalisis. Dari semua data yang ada yang digunakan hanya pekerjaan, usia pernikahan, jumlah anggota keluarga, status kepemilikan tanah, atap, langit-langit, lantai, dinding. Karena informasi yang ada didalamnya sudah mewakili informasi yang dibutuhkan untuk dijadikan indikator penelitian.

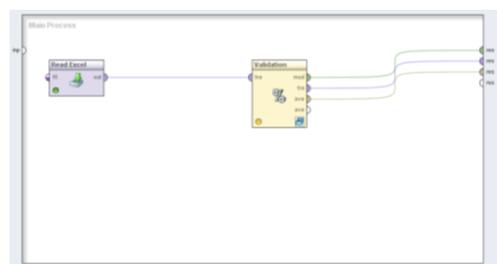
Hasil kriteria yang digunakan adalah pekerjaan, usia pernikahan, jumlah anggota keluarga, status kepemilikan tanah, atap, langit-langit, lantai, dinding. Dari 8 kriteria tersebut sangat mempengaruhi hasil sehingga mendapatkan akurasi yang baik.

c. Proses Training dan Testing

Sebelum dilakukan proses *training* perlu menentukan nilai-nilai parameter, jumlah data *training*, kelas data *training* dan data testing sebgailangkah awal penyimpanan vektor-vektor atribut dan kelas klasifikasi dari data. Tujuan dari proses *testing* adalah mendapatkan *output* target kelas yang hasilnya dicocokkan dengan kelas klasifikasi pada data *testing* untuk mengetahui presentasi keakurasiannya. Pada penelitian ini pengujian Algoritma menggunakan algoritma *Naive Bayes* dan *C.45* pada penelitian ini penulis membuat tabel data *training* yang digunakan dalam penelitian

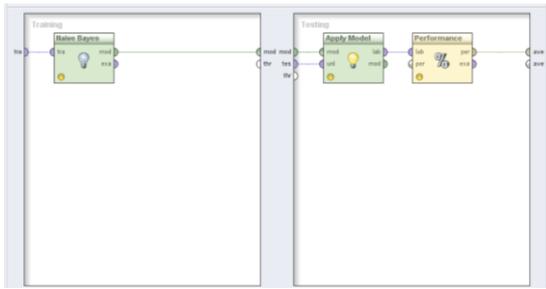
d. Perhitungan Algoritma Naive Bayes menggunakan RapidMiner

Setelah dilakukan pengujian manual maka akan dilanjutkan dengan pengujian *software RapidMiner* dengan tujuan untuk mendapatkan informasi pengujian algoritma yang dilihat dari nilai akurasi yang akan ditampilkan oleh *RapidMiner*. Pengujian ini menggunakan 114 *record* data. Keakurasian dari hasil uji coba menggunakan algoritma *Naive Bayes* ditinjau dari dua parameter yaitu *x-validation* dan jumlah data *training*. Uji coba algoritma *Naive Bayes* dilakukan dengan cara menentukan nilai *x-validation* dan jumlah data *training* yang digunakan. Hasil akurasi klasifikasi yang ditinjau dari parameter = 2, 3, 4, 5,..... 10 sebagai berikut:



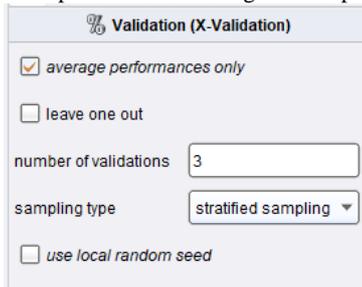
Gambar 1 Proses Analisis Data Dengan Tools RapidMiner

Pada gambar diatas di awali dengan operator *read excel* dan *X-Validation*, pada operator *read excel* mengimputkan data *training* dan menghubungkan operator *read excel* dengan operator *X-Validation*.



Gambar 2. Analisis Algoritma Naive Bayes

Dari Gambar 2 menjelaskan operator Naive Bayes, Apply Model dan Performance yang terdapat pada operator X-Validation. Kemudian semua operator itu dihubungkan satu persatu.



Gambar 3. Number Of Validation

Pada Gambar 4. operator X-Validation menggunakan parameter 2-10 dengan menggunakan sampling type stratified sampling untuk 10 kali pengujian.

	true TIDAK LAYAK	true LAYAK	class precision
pred. TIDAK LAYAK	108	4	96.43%
pred. LAYAK	1	1	50.00%
class recall	99.08%	20.00%	

Gambar 4. Akurasi Naive Bayes

Pada gambar 4. menjelaskan proses terakhir dalam pengujian mencari nilai akurasi tertinggi. Dengan melakukan Pengujian menggunakan X-Validation dan sampling type stratified sampling dengan sepuluh kali pengujian, dengan mengganti nilai number of validation 2-10 hingga diperoleh nilai Akurasi tertinggi.

Tabel 1 Pengujian tingkat Akurasi Algoritma Naive Bayes

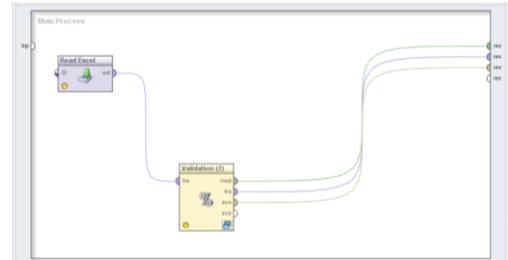
X-Validation	Accuracy	class precision	class recall
2	93.86%	0.00%	0.00%
3	95.61%	50.00%	16.67%
4	93.90%	0.00%	0.00%
5	94.74%	33.33%	20.00%
6	94.74%	0.00%	0.00%
7	94.75%	33.33%	20.00%
8	93.81%	0.00%	0.00%
9	95.58%	50.00%	20.00%
10	93.86%	0.00%	0.00%

Berdasarkan tabel 1. terlihat bahwa percobaan algoritma Naive Bayes dengan 10 variabel periode dengan menggunakan cross validation dan mengganti nilai input number of validation = 2 hingga 10. Dari hasil percobaan yang dilakukan dengan algoritma Naive Bayes, maka diperoleh akurasi tertinggi pada x validation = 3 accuracy 95.61% +/- 1.24% (mikro: 95.61%) class precision 50.00% dan class recall 16.67%.

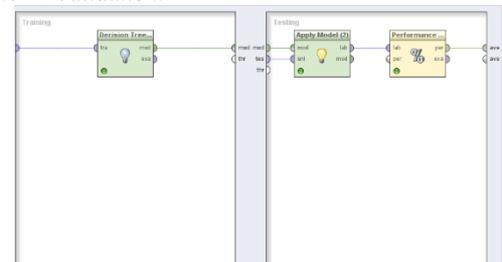
e. Pengujian Algoritma C4.5 Menggunakan RapidMiner

Setelah dilakukan pengujian manual maka akan dilanjutkan dengan pengujian software RapidMiner dengan tujuan untuk mendapatkan informasi pengujian algoritma yang dilihat dari

nilai akurasi yang akan ditampilkan oleh RapidMiner. Pengujian ini menggunakan 114 record data, keakurasian dari hasil uji coba menggunakan algoritma Naive Bayes ditinjau dari dua parameter yaitu x validation dan jumlah data training. Uji coba algoritma Naive Bayes dilakukan dengan cara menentukan nilai x-validation dan jumlah data training yang digunakan. Hasil akurasi klasifikasi yang ditinjau dari parameter = 2, 3, 4, 5,..... 10 sebagai berikut:

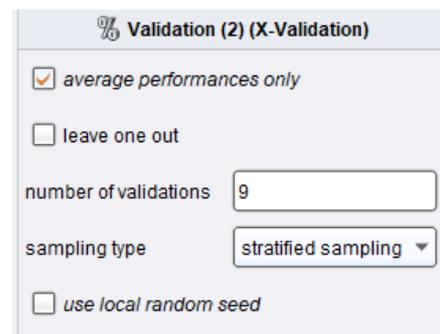


Gambar 5. Proses Analisis Data Dengan Tools RapidMiner Pada gambar 5 diawali dengan operator read excel dan X-Validation, pada operator read excel mengimputkan data training dan menghubungkan operator read excel dengan operator X-Validation.



Gambar 6. Analisis Dengan Algoritma C.45

Dari Gambar menjelaskan operator Decision Tree, Apply Model dan Performance yang terdapat pada operator X-Validation. Untuk operator C4.5 mengganti criterion gain ratio menjadi accuracy kemudian semua operator itu dihubungkan satu persatu.



Gambar 7. Tingkat Akurasi C4.5

Pada Gambar 7 operator X-Validation menggunakan parameter 2-10 dengan menggunakan sampling type stratified sampling untuk 10 kali pengujian.

	true TIDAK LAYAK	true LAYAK	class precision
pred. TIDAK LAYAK	109	4	96.46%
pred. LAYAK	0	1	100.00%
class recall	100.00%	20.00%	

Gambar 8. Number Of Validation

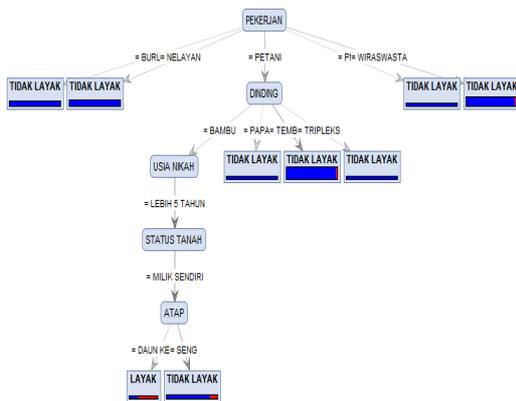
Dengan melakukan Pengujian menggunakan X-Validation dan sampling type stratified sampling dengan sepuluh kali pengujian, dengan mengganti nilai number of validation 2-10 hingga diperoleh nilai Akurasi tertinggi. Hasil pengujian tingkat akurasi dapat dilihat pada tabel 2.

Tabel 2. Pengujian tingkat Akurasi Algoritma C4.5

X-Validi tion	Accuracy	class precision	class recall
2	94.74%	0.00%	0.00%
3	96.49%	100.00%	16.67%
4	95.66%	50.00%	25.00%
5	95.61%	0.00%	0.00%
6	95.61%	50.00%	20.00%
7	96.43%	100.00%	20.00%
8	95.54%	50.00%	20.00%
9	96.51%	100.00%	20.00%
10	95.61%	50.00%	20.00%

Berdasarkan tabel 2. terlihat bahwa percobaan dengan 10 variabel periode dengan menggunakan *cross validation* dan mengganti nilai *input number of validation* = 2-10. Dari hasil percobaan yang dilakukan dengan algoritma C4.5, maka diperoleh nilai tertinggi pada *x validation* = 9 *accuracy* 96.51% +/- 3.91% (mikro: 96.49%) *class precision* 100.00% dan *class recall* 20.00%.

Untuk hasil pohon keputusan dapat dilihat bahwa *node* tertinggi adalah pekerjaan dengan demikian kondisi dari *node* tertinggi bisa di jadikan sebagai faktor yang paling menentukan dalam pemberian bantuan rumah sehat. Pohon keputusan yang terbentuk pada gambar 10 dan *rule* dari *decision tree*.



Gambar 9. Pohon Keputusan Node Tertinggi

```

Tree
PEKERJAAN = BURUH: TIDAK LAYAK (TIDAK LAYAK=9, LAYAK=0)
PEKERJAAN = NELAYAN: TIDAK LAYAK (TIDAK LAYAK=14, LAYAK=0)
PEKERJAAN = PETANI
|
| DINDING = BAMBUI
| |
| | USIA NIKAH = LEBIH 5 TAHUN
| | |
| | | STATUS TANAH = MILIK SENDIRI
| | | |
| | | | ATAP = DAUN KELAPA: LAYAK (TIDAK LAYAK=1, LAYAK=2)
| | | | |
| | | | | ATAP = SENG: TIDAK LAYAK (TIDAK LAYAK=6, LAYAK=1)
| | | | | |
| | | | | | DINDING = PAPAN: TIDAK LAYAK (TIDAK LAYAK=3, LAYAK=0)
| | | | | | |
| | | | | | | DINDING = TEMBOK: TIDAK LAYAK (TIDAK LAYAK=42, LAYAK=1)
| | | | | | | |
| | | | | | | | DINDING = TRIPLEKS: TIDAK LAYAK (TIDAK LAYAK=4, LAYAK=0)
| | | | | | | | PEKERJAAN = PNS: TIDAK LAYAK (TIDAK LAYAK=3, LAYAK=0)
| | | | | | | | PEKERJAAN = WIRASWASTA: TIDAK LAYAK (TIDAK LAYAK=27, LAYAK=1)
    
```

Gambar 9. Rule Decision Tree

Keterangan:

Pohon keputusan tersebut, bisa dilihat bahwa *node* tertinggi adalah pekerjaan, dengan demikian kondisi pekerjaan dijadikan sebagai faktor yang paling menentukan dalam penentuan penerima bantuan rumah sehat.

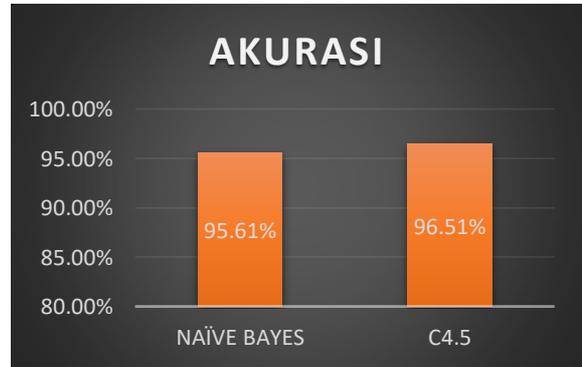
Dari kedua pengujian yang dilakukan dengan algoritma *Naive Bayes* dan C4.5 maka diperoleh nilai *accuracy* terbaik terdapat pada Algoritma C4.5 dengan nilai akurasi 96.51%, *class precision* 100.00% dan *class recall* 20.00%

f. Analisis Perbandingan Algoritma Naive Bayes dan C4.5

Dalam perbandingan kedua algoritma, proses *training* menggunakan parameter yang sama. Hal ini terlihat pada

Pengujian menggunakan *X-Validation* dan *sampling type stratified sampling* dengan sepuluh kali pengujian dengan mengganti nilai *number of validation* 2-10 hingga diperoleh nilai Akurasi tertinggi. Dari hasil uji coba, perbandingan dua metode ini didasarkan pada hasil klasifikasi dengan jumlah data training yang digunakan dan untuk selanjutnya diambil nilai rata-rata akurasi tertinggi pada setiap parameter yang ditentukan guna memperoleh hasil perbandingan dari kedua metode.

Berdasarkan analisis data yang telah dilakukan maka didapatkan tingkat akurasi dari kedua algoritma dapat dilihat pada gambar 12 berikut:



Gambar 10. Hasil Perbandingan Algoritma

Dari gambar diatas dapat dilihat bahwa perbandingan rata-rata tingkat akurasi hasil klasifikasi menggunakan algoritma *Naive Bayes* dan C4.5 Hasil uji coba dengan algoritma *Naive Bayes* menunjukkan bahwa dari keseluruhan data, diperoleh hasil sehingga nilai akurasinya sebesar 95.61 %. Sedangkan uji coba dengan algoritma C.45 menunjukkan bahwa dari keseluruhan data, diperoleh hasil sehingga nilai akurasinya sebesar 96.51 %. Jika ditinjau dari performansi dalam melakukan proses *running* program, pada gambar 8 terlihat algoritma C4.5 mendapatkan nilai akurasi yang tertinggi dibandingkan dengan algoritma *Naive Bayes* dalam pengklasifikasian data bantuan rumah sehat di Desa Dambalo Kecamatan Tomilito, Kabupaten Gorontalo Utara.

g. Hasil Klasifikasi

Pada tahapan ini peneliti menggunakan data *training* sebanyak 100 data dan data *testing* sebanyak 14 dengan *X-Validation* = 9, maka hasil klasifikasi yang diprediksi tidak layak berjumlah 13 data dan yang layak berjumlah 1 data, maka hasil klasifikasi menggunakan kedua algoritma untuk klasifikasi bantuan rumah sehat dapat dilihat pada gambar 13 dibawah ini:

Row No.	KATEGORI	confidence	precision	PEKERJAAN	USIA NIKAH	JUMLAH AN	STATUS TA	ATAP	LANGIT-LA	LANTAI	DINDING	
1	?	1	0	TIDAK LAYAK BURUH	LEBIH 5 TA	LEBIH 2 OR	MILIK SEND	SENG	KAYU	LANTAI SEM	BAMBUI	
2	?	0.975	0.025	TIDAK LAYAK PETANI	LEBIH 5 TA	LEBIH 2 OR	MILIK SEND	GENTENGS	TRIPLEKS	TEGEL	TEMBOK	
3	?	1	0	TIDAK LAYAK NELAYAN	LEBIH 5 TA	KURANG 2	MILIK SEND	SENG	KAYU	TEGEL	TEMBOK	
4	?	1	0	TIDAK LAYAK NELAYAN	LEBIH 5 TA	LEBIH 2 OR	MILIK SEND	SENG	KAYU	LANTAI SEM	BAMBUI	
5	?	0.975	0.025	TIDAK LAYAK PETANI	LEBIH 5 TA	LEBIH 2 OR	MILIK SEND	GENTENGS	TRIPLEKS	LANTAI SEM	TEMBOK	
6	?	0.800	0.200	TIDAK LAYAK PETANI	LEBIH 5 TA	LEBIH 2 OR	MILIK SEND	SENG	KAYU	LANTAI SEM	BAMBUI	
7	?	0.975	0.025	TIDAK LAYAK PETANI	LEBIH 5 TA	KURANG 2	MILIK SEND	SENG	TRIPLEKS	TEGEL	TEMBOK	
8	?	1	0	TIDAK LAYAK NELAYAN	LEBIH 5 TA	LEBIH 2 OR	MILIK SEND	SENG	KAYU	LANTAI SEM	BAMBUI	
9	?	0.957	0.043	TIDAK LAYAK WIRASWASTA	LEBIH 5 TA	LEBIH 2 OR	BUKAN MILIK	SENG	BAMBUI	LANTAI SEM	TEMBOK	
10	?	0.333	0.667	LAYAK	PETANI	LEBIH 5 TA	LEBIH 2 OR	MILIK SEND	DAUN KELPA	KAYU	LANTAI SEM	BAMBUI
11	?	0.975	0.025	TIDAK LAYAK PETANI	LEBIH 5 TA	LEBIH 2 OR	MILIK SEND	GENTENGS	TRIPLEKS	TEGEL	TEMBOK	
12	?	0.975	0.025	TIDAK LAYAK PETANI	LEBIH 5 TA	LEBIH 2 OR	MILIK SEND	SENG	TRIPLEKS	LANTAI SEM	TEMBOK	
13	?	0.800	0.200	TIDAK LAYAK PETANI	LEBIH 5 TA	LEBIH 2 OR	MILIK SEND	SENG	KAYU	LANTAI SEM	BAMBUI	
14	?	0.957	0.043	TIDAK LAYAK WIRASWASTA	LEBIH 5 TA	KURANG 2	MILIK SEND	SENG	TRIPLEKS	LANTAI SEM	TEMBOK	

Gambar 12. Hasil Klasifikasi

4. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan maka dapat diambil kesimpulan diantaranya:

1. Dengan menganalisa hasil perhitungan perbandingan dengan menggunakan algoritma *Naive Bayes* dan C4.5 maka dari pengujian hasil klasifikasi bantuan rumah sehat menggunakan 114 data sampel yang telah dilakukan diperoleh nilai akurasi yang berbeda jauh, tingkat akurasi *naive bayes* sebesar 95.61 % sedangkan tingkat akurasi algoritma C4.5 sebesar 96.51 %.

2. Nilai keakurasian tertinggi hasil klasifikasi yang ditinjau dari parameter algoritma C4.5 mendapatkan nilai akurasi yang tertinggi dibandingkan dengan algoritma *Naïve Bayes*.
3. Dari kedua hasil akurasi berhasil divalidasi dan di implementasi dapat mengklasifikasikan jumlah penerima bantuan rumah sehat yang telah menjadi program pemerintah pusat. Berdasarkan data *training* dan pengujian langkah berikutnya melakukan klasifikasi sehingga menghasilkan kesimpulan bahwa jumlah data *training* mempengaruhi pada presentase keakurasian.
4. Penerapan Algoritma *Naïve Bayes* dan C4.5 untuk mengklasifikasi data calon penerima bantuan rumah sehat mendapatkan hasil akurasi yang belum memuaskan di karenakan dengan jumlah *record* data yang sedikit.

Rumah di Kelurahan KRrapyak. Sari & Prabowo, 148, 148–162. [http://repository.untag-sby.ac.id/5231/49/BAB II.pdf](http://repository.untag-sby.ac.id/5231/49/BAB%20II.pdf)

- [12] Utomo, D. P., & Mesran, M. (2020). Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung. *Jurnal Media Informatika Budidarma*, 4(2), 437. <https://doi.org/10.30865/mib.v4i2.2080>

DAFTAR PUSTAKA

- [1] Aditia. (2013). Hubungan Cemaran Mikroba Dengan Pengelolaan Rumah sehat Pada Rumah. *Journal of Chemical Information and Modeling*, 53(9), 1689–1699. <http://eprints.umm.ac.id/35046/3/jiptummpp-gdl-aditiahuda-47406-3-babii.pdf>
- [2] Elisa, E. (2017). Analisa dan Penerapan Algoritma C4.5 Dalam Data Mining Untuk Mengidentifikasi Faktor-Faktor Penyebab Kecelakaan Kerja Kontruksi PT.Arupadhatu Adisesanti. *Jurnal Online Informatika*, 2(1), 36. <https://doi.org/10.15575/join.v2i1.71>
- [3] Faiz Zamzami & Nabella Duta Nusa, 2017 : 32. (2020). Bab Ii Landasan Teori. *Riswanto & Laluma*, 53(9), 8–24.Hamsinar. (2019). *Surveilans Epidemiologi Program Studi SI Kesehatan Masyarakat Fakultas Kesehatan Dan Farmasi Universitas Muhammadiyah 2018 / 2019*. 28.
- [4] 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>
- [5] Kurniawan, Y. I. (2017). Perbandingan Algoritma *Naive Bayes* dan C.45 dalam Klasifikasi Data Mining. . . *Bansar, Sharma & Goel*, 5(4), 455. <https://doi.org/10.25126/jtiik.201854803>
- [6] Prapitasari, I., Utami, K., Setyarini, R., ... S. A.-J. K., 2017, undefined, Pandiana, L., Komunitas, S. N.-J. K., & 2018, undefined. (2019). *Bab Iii Metodologi Penelitian*. 62–76. <https://www.neliti.com/publications/275650/elderly-visit-routines-to-elderly-integrated-service-post-in-the-working-area-of%0Ahttps://ejournal3.undip.ac.id/index.php/jkm/article/view/16353%0Ahttp://repository.unjaya.ac.id/id/eprint/3305>
- [7] Manalu, E., Sianturi, F. A., & Manalu, M. R. (2017). Penerapan Algoritma *Naive Bayes* Untuk Memprediksi Jumlah Produksi Barang Berdasarkan Data Persediaan dan Jumlah Pemesanan Pada CV. Papadan Mama Pastries. *Jurnal Mantik Penusa*, 1(2), 16–21.
- [8] Nasrullah, A. H. (2018). Penerapan Metode C4.5 untuk Klasifikasi Mahasiswa Berpotensi Drop Out. *ILKOM Jurnal Ilmiah*, 10(2), 244–250. <https://doi.org/10.33096/ilkom.v10i2.300.244-250>
- [9] Rani, L. N. (2016). Klasifikasi Nasabah Menggunakan Algoritma C4.5 Sebagai Dasar Pemberian Kredit. *INOVTEK Polbeng - Seri Informatika*, 1 (2), 126. <https://doi.org/10.35314/isi.v1i2.131>
- [10] Sholeh, M. (2014). Prosiding Seminar Nasional Aplikasi Sains & Teknologi (SNAST) 2014 Yogyakarta, 15 November 2014 ISSN: 1979-911X. *Snast, November*, 211–216.
- [11] Sari & Prabowo. (2017). Bagaimana Menerapkan Algoritma *Naive Bayes* Untuk PEnentuan Bantuan Bedah