

Analisis Performa Algoritma Klasifikasi *Naive Bayes* dan C4.5 untuk Prediksi Penerima Bantuan Jaminan Kesehatan

Nurfazriah Attamami ¹, Agung Triayudi ^{2*}, Rima Tamara Aldisa ³

^{1,2,3} Fakultas Teknologi Komunikasi dan Informatika, Universitas Nasional, Kota Jakarta Selatan, Daerah Khusus Ibukota Jakarta, Indonesia.

article info

Article history:

Received 11 August 2022

Received in revised form

7 December 2022

Accepted 16 February 2023

Available *online* April 2023

DOI:

<https://doi.org/10.35870/jtik.v7i2.756>

Keywords:

Data Mining; Decision Tree; Naïve Bayes; Assistance Program; Health Insurance.

Kata Kunci:

Data Mining; Decision Tree; Naïve Bayes; Program bantuan; Jaminan Kesehatan.

abstract

Health is the basis of the level of humanity. However, in reality, not everyone with social welfare problems has national health insurance. The large number of patient files that must be checked makes it difficult for officers to identify potential beneficiaries. Based on these problems, a procedure or method is needed that can assist officers in identifying potential beneficiaries. From the results of the performance testing of the two models using a confusion matrix with 730 records used as training data and 313 records used as test data, the C4.5 classification algorithm gets the highest accuracy value, which is 99.04%. A total of 310 data records were predicted correctly with an error rate or error of 0.96% or as many as 3 data records from 313 data tested. While the Naive Bayes classification algorithm gets an accuracy value of 92.97%. A total of 291 data records were predicted to be correct with an error rate of 7.03% or as many as 22 data records were predicted to be incorrect from the 313 data tested.

abstract

Kesehatan adalah dasar dari taraf kemanusiaan. Namun kenyataannya, tidak semua orang dengan masalah kesejahteraan sosial memiliki jaminan kesehatan nasional. Banyaknya berkas pasien yang harus diperiksa menyulitkan petugas untuk mengidentifikasi calon penerima bantuan. Berdasarkan permasalahan tersebut, diperlukan suatu prosedur atau metode yang dapat membantu petugas dalam mengidentifikasi calon penerima bantuan. Dari hasil pengujian kinerja kedua model menggunakan confusion matrix dengan 730 record yang digunakan sebagai data latih dan 313 record yang digunakan sebagai data uji, algoritma klasifikasi C4.5 mendapatkan nilai akurasi yang paling tinggi yaitu sebesar 99.04%. Sebanyak 310 record data diprediksi tepat dengan tingkat error atau kesalahan sebesar 0.96% atau sebanyak 3 record data dari 313 data yang diuji. Sedangkan pada algoritma klasifikasi Naive Bayes di dapatkan nilai akurasi sebesar 92.97%. Sebanyak 291 record data diprediksi tepat dengan tingkat error atau kesalahan sebesar 7.03% atau sebanyak 22 record data diprediksi salah dari 313 data yang diuji.

Corresponding Author. Email: agungtriayudi@civitas.unas.ac.id ^{2}.

1. Latar Belakang

Kesehatan adalah dasar dari taraf kemanusiaan dan penghormatan terhadap hak asasi manusia [1]. Setiap insan berhak memiliki kehidupan yang sehat dan berhak atas perawatan medis yang baik. BPJS PBI merupakan salah satu jenis kepesertaan yang ditujukan untuk fakir miskin sesuai kriteria yang telah diatur kementerian sosial dengan iuran setiap bulannya ditanggung dan dibayarkan oleh pemerintah [2]. Dalam pelaksanaannya tidak semua masyarakat yang menyandang masalah kesejahteraan sosial terdaftar menjadi peserta BPJS PBI sehingga masih terdapat masyarakat yang belum tercover jaminan kesehatan nasional. Berdasarkan Perpres Nomor 111 Tahun 2013 menyatakan bahwa masyarakat miskin atau orang tidak mampu yang tidak terdaftar kepesertaan BPJS PBI dapat diberikan bantuan jaminan kesehatan oleh pemerintah [3]. Sesuai yang tercantum pada Perwal Kota Depok No. 4 Tahun 2016, Pemerintah Kota Depok menyediakan jaminan pembiayaan kesehatan bagi masyarakat yang memiliki masalah kesejahteraan sosial yang tidak terdaftar dalam kepesertaan BPJS PBI melalui kriteria selektif dan memenuhi persyaratan penerima bantuan [4].

Dalam prosesnya, seleksi calon penerima bantuan jaminan kesehatan memiliki kendala dalam mengidentifikasi calon penerima bantuan. Banyaknya berkas-berkas pasien yang harus diperiksa membuat petugas kesulitan dalam mengidentifikasi calon penerima bantuan jaminan kesehatan dan terkadang proses identifikasi menjadi kurang tepat. Metode yang dapat dilakukan untuk membantu mengatasi permasalahan ini diantaranya yaitu dengan melakukan otomatisasi dan komputerisasi prosedur yang sudah berjalan saat ini, salah satunya dengan menggunakan sistem pembelajaran mesin dengan algoritma klasifikasi Data Mining.

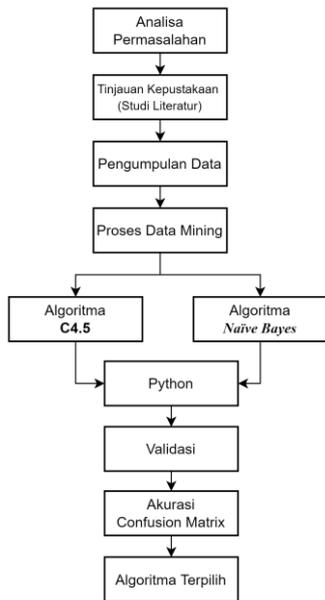
Penelitian Saputra, Wasiyanti, & Pribadi, (2021) melakukan klasifikasi data mining menggunakan algoritma klasifikasi C4.5 untuk menentukan kriteria kelayakan calon penerima BPNT. Hasil akhir dari penelitian ini menunjukkan model yang digunakan memiliki nilai akurasi yang cukup tinggi sebesar 91,54% [5]. Penelitian Sulihati (2022) melakukan komparasi dua algoritma data mining Naive Bayes dan Pohon Keputusan (Decision Tree) untuk

memprediksi seleksi penerimaan masuk perguruan tinggi negeri. Hasil dari penelitian dengan menggunakan algoritma pohon keputusan menghasilkan persentase akurasi sebesar 82,34%, sedangkan untuk algoritma Naive Bayes menghasilkan persentase akurasi sebesar 91,08% [6]. Penelitian Nul Hakim, Sholihin, & Rizki Rinaldi (2021) menggunakan algoritma naive bayes untuk memprediksi penerima program bantuan keluarga pra sejahtera. Dari hasil penelitian menghasilkan tingkat akurasi sangat baik sebesar 99,14 % [7]. Penelitian menggunakan algoritma klasifikasi C4.5 yang dilakukan oleh Ubaedi & Djaksana (2022) bertujuan untuk memprediksi kelayakan kredit. Dari hasil penelitian didapatkan persentase akurasi sebesar 79.11% [8]. Penelitian Bahtiar & Silitonga (2020) menerapkan algoritma klasifikasi C4.5 untuk memprediksi kriteria calon penerima PKH di Kecamatan Sumber Kabupaten Cirebon. Dari penelitian ini didapatkan persentase akurasi yang tinggi yaitu 98.30% [9].

Berdasarkan uraian permasalahan diatas, dapat diketahui bahwa penggunaan data mining dapat membantu pengklasifikasian data dengan cepat dan tingkat akurasi yang akurat sehingga sangat tepat digunakan untuk pengklasifikasian kelayakan calon penerima jaminan kesehatan. Dalam studi ini, peneliti akan menerapkan algoritma data mining Naive Bayes dan C4.5 untuk menentukan kelayakan calon penerima jaminan kesehatan bagi masyarakat yang memiliki masalah kesejahteraan sosial dan melakukan analisis dan perbandingan guna memahami kinerja kedua algoritma tersebut sehingga dapat dipilih algoritma yang memiliki tingkat akurasi terbaik. Dengan adanya penelitian ini diharapkan dapat membantu petugas dinas kesehatan dalam menentukan kelayakan calon penerima jaminan kesehatan bagi masyarakat yang memiliki masalah kesejahteraan sosial.

2. Metode Penelitian

Dalam langkah penelitian ini dibuat keseluruhan alur penelitian, dimulai dari tahap pengumpulan data hingga memperoleh hasil akhir yang didapat oleh peneliti. Adapun tahapan yang dilakukan dalam penelitian ini seperti pada gambar alur penelitian berikut:



Gambar 1. Alur Penelitian

Berdasarkan gambar diatas dapat dijelaskan sebagai berikut :

Analisa Permasalahan

Tahapan ini melakukan perumusan identifikasi masalah yang terjadi dan solusi yang diharapkan untuk permasalahan ini.

Tinjauan Kepustakaan

Tahapan ini penulis melakukan studi pustaka dan dokumen terkait dengan objek penelitian, serta melakukan wawancara dan observasi langsung ke Dinas Kesehatan Kota Depok untuk memperoleh data dan informasi yang diperlukan selama penelitian.

Pengumpulan Data

Tahapan ini penulis melakukan studi pustaka dan dokumen terkait dengan objek penelitian, serta melakukan wawancara dan observasi langsung ke Dinas Kesehatan Kota Depok untuk memperoleh data dan informasi yang diperlukan selama penelitian.

Proses Data Mining

Dalam tahapan ini proses pengolahan data dilakukan dengan metode standarisasi data mining yaitu CRISP–DM. Ada 6 tahapan yang dilakukan dalam metode ini yaitu; Tahapan Pemahaman Proses Bisnis, Tahapan Pendalaman Data, Tahapan Persiapan Data, Tahapan Pemodelan, Tahapan Evaluasi dan Tahapan Penyebaran [10].

a. Tahapan Pemahaman Proses Bisnis. Tahapan ini dilakukan untuk mengetahui tujuan dari penelitian

dan mencari solusi yang dapat diterapkan untuk menyelesaikan masalah dengan menggunakan data mining.

b. Tahapan Pendalaman Data. Tahapan ini terdiri dari pengumpulan data dan pendeskripsian data, pendalaman data dan verifikasi kualitas data[14]. Setelah data berhasil dikumpulkan kemudian dilakukan pemahaman data dengan mempelajari variabel yang digunakan. Terdapat 13 variabel yang digunakan sebagai atribut kriteria seperti yang ditunjukkan pada gambar berikut :

Gambar 2. Atribut Kriteria

c. Tahapan Persiapan Data. Fase persiapan data atau preprocessing data meliputi transformasi dan pembersihan data [12]. Pada fase ini data yang telah diterima kemudian dilakukan cleansing data atau penyeleksian data guna menghasilkan data yang dibutuhkan. Sebanyak 1325 data pemohon bantuan pembiayaan jaminan kesehatan dilakukan pembersihan data kemudian diseleksi agar menghasilkan data yang bersih sehingga data yang tersisa sebanyak 1043 data yang akan digunakan untuk fase pemodelan.

d. Tahapan Pemodelan Data. Pada tahap pemodelan, data yg telah diseleksi lalu dilakukan pemodelan menggunakan python dan diuji menggunakan Confusion Matrix dengan menerapkan algoritma Naive Bayes dan C4.5 Decision Tree. Algoritma C4.5 Decision Tree merupakan algoritma yang membentuk struktur pohon keputusan [15]. Sedangkan algoritma Naive Bayes merupakan algoritma yang menggunakan pengklasifikasian statistik untuk memprediksi probabilitas suatu kelas [16].

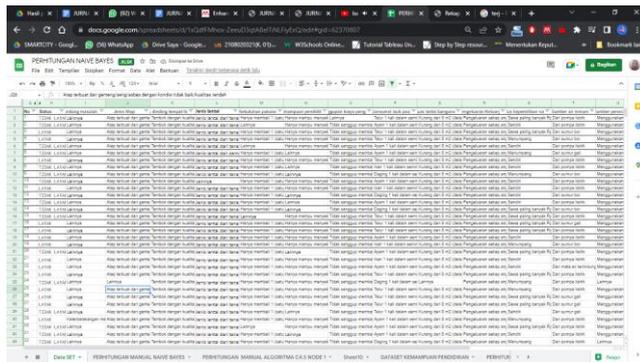
e. Tahapan Evaluasi. Tahapan ini dilakukan untuk menguji daya guna dan kualitas dari model yang digunakan. Hasil prediksi kemudian diuji dengan menggunakan confusion matrix untuk mengukur performa dari model klasifikasi yang sudah dibuat.

f. Tahapan Penyebaran. Pada fase ini setelah dilakukan komparasi dan evaluasi, algoritma dengan nilai tingkat akurasi terbesar akan dipilih dan diterapkan dengan menggunakan data baru untuk menghasilkan prediksi atau hasil yang diharapkan sehingga dapat memberikan interpretasi terhadap dataset dari masing-masing algoritma yang diterapkan.

3. Hasil dan Pembahasan

Sumber Data

Sumber data yg dipakai merupakan data permohonan bantuan pembiayaan jaminan kesehatan untuk fakir miskin menurut Dinas Kesehatan Kota Depok tahun 2020-2022 sebanyak 1043 record data yang sudah dilakukan pembersihan dan penyeleksian data. Dari dataset tersebut digunakan 730 dataset sebagai data pelatihan dan 313 dataset digunakan sebagai data pengujian. Data pelatihan digunakan sebagai data latih algoritma untuk mendapatkan model yang paling sesuai, dan data pengujian digunakan untuk menguji kinerja model yang dibuat selama fase pengujian [13]. Gambar 3 merupakan dataset yang digunakan untuk pengujian data.



Gambar 3. Dataset Pengujian

Algoritma C4.5

Proses perhitungan dilakukan secara semi-manual menggunakan algoritma klasifikasi C4.5 untuk menghasilkan aturan dan membentuk pohon keputusan (Decision Tree). Berikut adalah langkah yang dilakukan dalam perhitungan algoritma klasifikasi C4.5 :

a. Menghitung total permohonan dengan status “LAYAK” dan “TIDAK LAYAK”. Kemudian hitung nilai total entropi untuk semua dataset. Diketahui dari dataset pengujian berisi 1043

record data. Sebanyak 616 record data dengan kelas “LAYAK” dan 427 record data dengan kelas “TIDAK LAYAK”. Dari keterangan tersebut didapat nilai entropi total sebagai berikut :

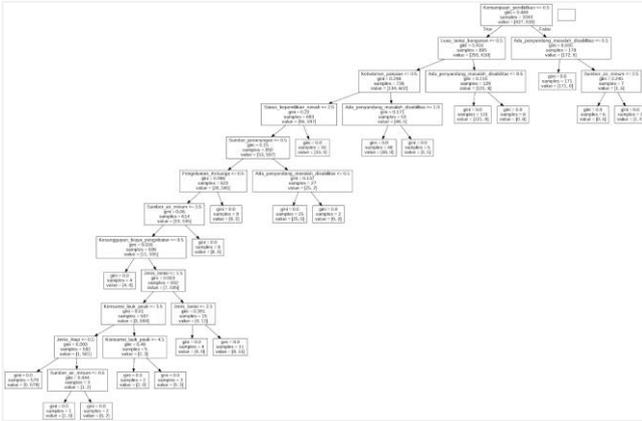
$$\begin{aligned}
 Entropy &= - \sum_{j=1}^m f(i,j) \cdot \log_2 f(i,j) \\
 &= (- 616/1043 \\
 &\quad * \log_2(616/1043)) \\
 &\quad + (-427/1043 \\
 &\quad * \log_2(427/1043)) \\
 &= 0,9762
 \end{aligned}$$

b. Setelah mendapatkan nilai entropi total, kemudian menghitung nilai entropi dan nilai gain dari setiap atribut. Hasil nilai gain tertinggi akan menjadi akar pohon (root). Gambar 4 merupakan hasil penghitungan nilai entropi dan gain pada simpul akar dari setiap atribut yang dihitung.

	Jumlah Kasus	LAYAK	TIDAK LAYAK	Entropi	Gain
Total	1043	616	427	0,9762	
Luas lantai bangunan					0,1287
perenang	914	608	306	0,9197	
Lainnya	129	8	121	0,3354	
Jenis lantai					0,0124
baik/kuualitas rendah	11	10	1	0,4395	
baik/kuualitas rendah	11	11	0	0	
baik/kuualitas rendah	109	74	35	0,9056	
baik/kuualitas rendah	918	531	387	0,9822	
Lainnya	4	0	4	0	
tinggi					0,0087
tembok yg sudah isang/berhant atau	12	9	3	0,8113	
tembok yg sudah isang/berhant atau	17	16	1	0,2228	
tembok yg sudah isang/berhant atau	1014	591	423	0,9801	
Jenis Atap					0,0046
tidak baik/kuualitas rendah	8	7	1	0,5436	
dengan kondisi tidak baik/kuualitas rendah	1033	609	424	0,9787	
Lainnya	2	0	2	0	
Sumber penoranngan					0,0266
(sembilan ratus) watt	1013	614	399	0,9673	
Lainnya	30	2	28	0,3534	
Sumber air minum					0,0177
Dari sumur bor	181	99	82	0,9936	
Dari sumur gali	102	66	33	0,9062	
Dari pompa listrik	707	415	292	0,9781	
Dari mata air terlindungi	44	33	11	0,8113	
Lainnya	9	0	9	0	
Konsumsi listrik peak					0,0613
Daging 1 kali dalam seminggu	130	48	82	0,9501	
Sosis 1 kali dalam seminggu	4	3	1	0,8113	
Ikan 1 kali dalam seminggu	204	101	103	0,9999	
Ayam 1 kali dalam seminggu	292	155	137	0,9973	
Telur 1 kali dalam seminggu	411	309	102	0,8084	
Lainnya	2	0	2	0	
Kebudayaan pakain					0,042
dalam 1 (satu) tahun	900	611	279	0,96	
Lainnya	53	5	48	0,4508	
pengobatan					0,0087
dipnakemas/poklain/RS	1036	616	420	0,974	
Lainnya	7	0	7	0	
Kemampuan pendidikan					0,2144
sampai jenjang pendidikan Sekolah Lanjutan	605	610	255	0,8748	
Lainnya	178	6	172	0,2127	
Pengeluaran Keluarga					0,0112
dari Rp. 650.000/kapita/bulan	1034	616	418	0,9734	
Lainnya	9	0	9	0	
Status kepemilikan rumah					0,0537
Sendiri	348	183	165	0,9981	
Menumpang	257	226	131	0,9483	
(tima ratus ribu per/bulan)	305	207	98	0,9098	
Lainnya	33	0	33	0	
Disabilitas					0,0333
Tuli	3	3	0	0	
Cacat fisik	12	11	1	0,4138	
Keterbelakangan mental	35	35	0	0	
Lainnya	993	567	426	0,9854	

Gambar 4. Nilai Entropi dan Gain pada Node Akar

c. Setelah node akar terbentuk, ulangi proses dengan cara yang sama pada semua kasus di setiap cabang sehingga semua cabang memiliki kelas yang sama. Gambar 5 adalah bentuk hirarki pohon keputusan dari hasil perhitungan yang sudah dilakukan.



Gambar 5. Hirarki Pohon Keputusan

- d. Berdasarkan gambar pohon keputusan atas, didapatkan aturan untuk memprediksi kelayakan calon penerima bantuan jaminan kesehatan. Aturan ini nantinya akan digunakan untuk memprediksi data baru.

Algoritma Naive Bayes

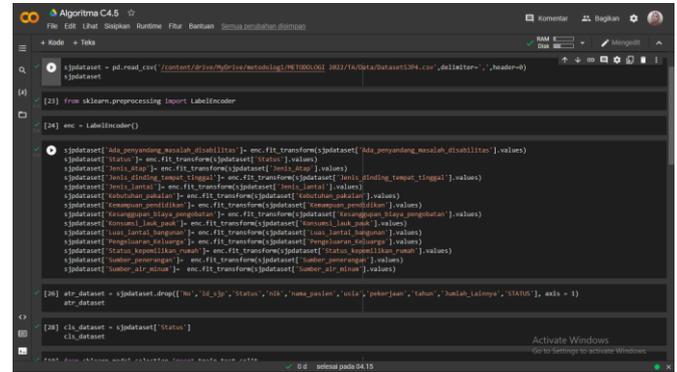
Langkah awal yang dilakukan dalam proses perhitungan algoritma naive bayes adalah mencari probabilitas dari setiap kelas yaitu “LAYAK” dan “TIDAK LAYAK” dari setiap atribut [17]. Gambar 6 merupakan hasil perhitungan probabilitas hipotesis dari setiap kelas.

		LAYAK	TIDAK LAYAK
Luas lantai bangunan	Kurang dari 8 m ² (delapan meter persegi)	0.58	0.7166
	Lainnya	0.013	0.2834
Jenis lantai	Jenis lantai dari tanah dengan kondisi tidak	0.0162	0.0023
	Jenis lantai dari kayu dengan kondisi tidak	0.0016	0
	Jenis lantai dari semen dengan kondisi tidak	0.1201	0.082
	Jenis lantai dari tanah keramik kondisi tidak Lainnya	0.862	0.9063
Jenis dinding tempat tinggal	Bambu dengan kualitas rendah termasuk	0.0144	0.007
	Kayu dengan kualitas rendah termasuk	0.026	0.0023
	Fenbok dengan kualitas rendah termasuk	0.9594	0.9906
Jenis Atap	Atap terbuat dari ijuk/rumah dengan kondisi	0.0114	0.0023
	Atap terbuat dari genteng/wing/wales dengan	0.9384	0.993
	Lainnya	0	0.0049
Sumber penerangan	Menggunakan listrik paling tinggi 500	0.9968	0.9314
	Lainnya	0.0032	0.0656
Sumber air minum	Dari sumur bor	0.160	0.192
	Dari sumur gali	0.112	0.0773
	Dari pompa listrik	0.673	0.6838
	Dari mata air terlindungi Lainnya	0.0536	0.0426
Kontsumsi buk parkir	Daging 1 kali dalam seminggu	0.079	0.192
	Susu 1 kali dalam seminggu	0.0049	0.0023
	Ikan 1 kali dalam seminggu	0.16	0.2312
	Ayam 1 kali dalam seminggu Telur 1 kali dalam seminggu Lainnya	0.2516	0.3208
Kebudayaan pakain	Hanya membeli 1 (satu) pasang pakaian	0.9919	0.8876
	Lainnya	0.0081	0.1124
Kemampuan biaya pengeluaran	Tidak sanggup membayar biaya pengobatan	1	0.9836
	Lainnya	0	0.0164
Kemampuan pendidikan	Hanya mampu menyelesaikan anaknya	0.9903	0.5972
	Lainnya	0.0097	0.4028
Pengeharian kesehatan	Pengeharian setiap anggota keluarga kurang	1	0.9789
	Lainnya	0	0.0211
Status kepemilikan rumah	Sendiri	0.2971	0.3864
	Merempang	0.3669	0.3068
	Sewa paling banyak Rp. 500.000 (lima ratus ribu) Lainnya	0.336	0.2208
Tul	Tul	0.0049	0
	Cerat baik	0.0179	0.0023
	Keterbelakangan mental	0.0168	0
	Lainnya	0.9205	0.9977

Gambar 6. Hasil Perhitungan Probabilitas Kelas Algoritma Naive Bayes

Hasil Pengujian

Pemodelan data dilakukan dengan bahasa pemrograman Python dan library sklearn untuk menerapkan kedua algoritma dalam Python dan menggunakan Tools Google Colab untuk menjalankan model yang sudah dibuat.



Gambar 7. Tampilan Awal Pemrograman Python

Seperti yang ditunjukkan gambar 7. Sebelum proses pengujian, dataset yang ada dilakukan proses konversi nilai non numerik menjadi nilai numerik agar dataset dapat dimengerti oleh komputer. Selain itu supaya hanya kolom atribut yg diproses, maka kolom yang berfungsi sebagai label seperti Nama, Usia, Pekerjaan sementara di hapus. Selanjutnya dilakukan proses pengujian menggunakan algoritma klasifikasi C4.5 dan Naive Bayes dengan menggunakan data testing sebanyak 313 record data. Kemudian data hasil prediksi dievaluasi dengan menggunakan Confusion Matrix.

```

[28] cls_dataset = sjpdataset['Status']
cls_dataset

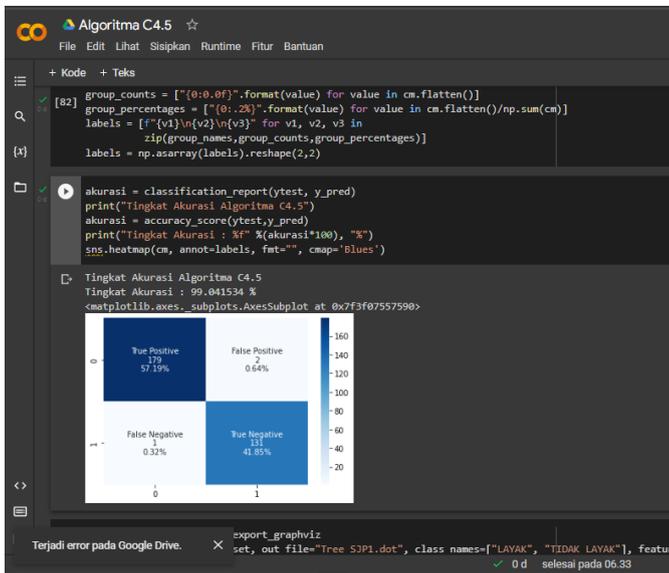
[29] from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score, confusion_matrix, classification_report
from sklearn.tree import DecisionTreeClassifier

xtrain, xtest, ytrain, ytest = train_test_split(atr_dataset, cls_dataset, test_size=0.3, random_state=1)
tree_dataset = DecisionTreeClassifier(random_state=1)
tree_dataset.fit(xtrain, ytrain)

[34] y_pred = tree_dataset.predict(xtest)
cm = confusion_matrix(ytest, y_pred)
print("Confusion Matrix")
print(cm)
akurasi = classification_report(ytest, y_pred)
print("Tingkat Akurasi Algoritma C4.5")
print("Akurasi : ", akurasi)
akurasi = accuracy_score(ytest, y_pred)
print("Tingkat Akurasi : %f" % (akurasi*100))

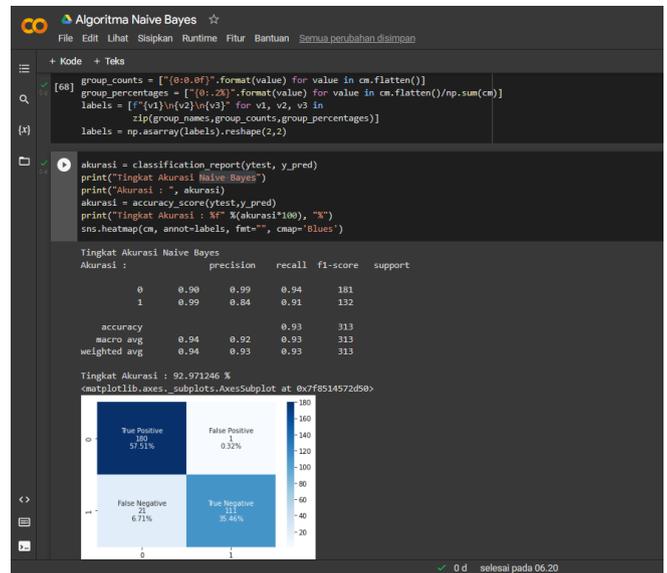
```

Gambar 8. Proses Pengujian Algoritma Klasifikasi C4.5



Gambar 9. Evaluasi Confusion Matrix C4.5

Dari gambar 9 didapatkan hasil tingkat akurasi dari algoritma C4.5 adalah 0.9904 atau 99.04%. Sebanyak 310 record data diprediksi benar dari total data pengujian yang dilakukan. Nilai Presisi sebesar 0.5774 atau 57.74%, nilai Recall sebesar 0.9944 atau 99.44%, nilai Specificity sebesar 0.9850 atau 98.50% dan F1 Score sebesar 0.7306 atau sebesar 73.06%. Dari tabel confusion matrix baris pertama dapat disimpulkan bahwa sebanyak 179 record data diprediksi benar sebagai kelas “LAYAK” dan 2 data diprediksi salah sebagai kelas “LAYAK”. Sedangkan pada baris kedua tabel confusion matrix dapat disimpulkan bahwa sebanyak 131 record data diprediksi benar sebagai kelas “TIDAK LAYAK” dan 1 data diprediksi salah sebagai kelas “TIDAK LAYAK”.



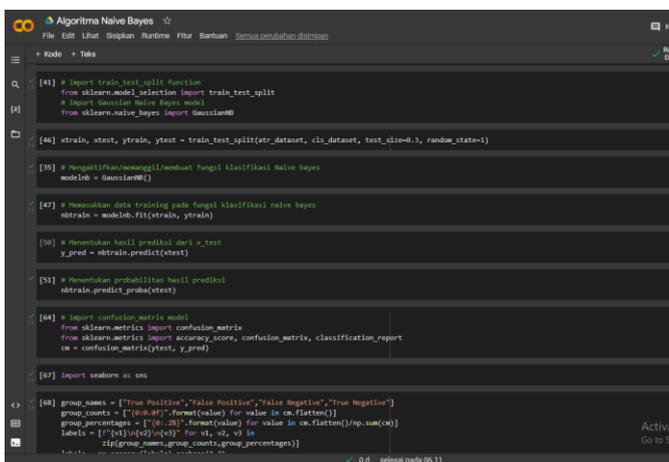
Gambar 11. Hasil Evaluasi Confusion Matrix Naïve Bayes

Berdasarkan gambar 11 didapatkan hasil tingkat akurasi dari algoritma Naïve Bayes adalah 0.9297 atau 92.97%. Sebanyak 291 dataset diprediksi dengan benar dari total data uji yang dilakukan. Nilai Presisi sebesar 0.6186 atau 61.86%, nilai Recall sebesar 0.8955 atau 89.55%, nilai Specificity sebesar 0.9911 atau 99.11% dan F1 Score sebesar 0.7317 atau sebesar 73.17%. Dari tabel confusion matrix baris pertama dapat disimpulkan bahwa sebanyak 180 record data diprediksi benar sebagai kelas “LAYAK” dan 1 data diprediksi salah sebagai kelas “LAYAK”. Sedangkan pada baris kedua tabel confusion matrix dapat disimpulkan bahwa sebanyak 111 record data diprediksi benar sebagai kelas “TIDAK LAYAK” dan 21 data diprediksi salah sebagai kelas “TIDAK LAYAK”.

Berdasarkan hasil pengujian yang telah dilakukan, perbandingan performa dari kedua algoritma yang diuji dapat dilihat pada tabel 1.

Tabel 1. Perbandingan Performa Algoritma C4.5 dan Naive Bayes

Uraian	C4.5	Naive Bayes
Akurasi	99.04%	92.97%
Presisi	57.74%	61.86%
Recall	99.44%	89.55%
Specificity	98.50%	99.11%
F1 Score	73.06%	73.17%



Gambar 10. Proses Pengujian Algoritma Naïve Bayes

4. Kesimpulan

Penelitian ini bertujuan untuk menganalisa performa dari algoritma klasifikasi Naive Bayes dan C4.5 terhadap dataset pemohon bantuan jaminan kesehatan. Berdasarkan hasil uji performa kedua model menggunakan metode evaluasi confusion matrix. Algoritma Klasifikasi C4.5 mendapatkan nilai akurasi yang paling tinggi yaitu sebesar 99.04%. Sebanyak 310 record data diprediksi benar dengan tingkat error atau kesalahan sebesar 0.96% atau sebanyak 3 record data dari 313 data yang diuji. Nilai Presisi sebesar 0.5774 atau 57.74%, nilai Recall sebesar 0.9944 atau 99.44%, nilai Specificity sebesar 0.9850 atau 98.50% dan F1 Score sebesar 0.7306 atau sebesar 73.06%. Sedangkan untuk algoritma Naive Bayes mendapatkan nilai akurasi sebesar 92.97%. Sebanyak 291 record data diprediksi benar dengan tingkat error atau kesalahan sebesar 7.03% atau sebanyak 22 record data diprediksi salah dari 313 data yang diuji. Nilai Presisi sebesar 0.5774 atau 57.74%, nilai Recall sebesar 0.9944 atau 99.44%, nilai Specificity sebesar 0.9850 atau 98.50% dan F1 Score sebesar 0.7306 atau sebesar 73.06%. Berdasarkan hasil diatas, dapat disimpulkan bahwa kedua algoritma yang diuji memiliki nilai akurasi yang cukup baik yaitu diatas 90%. Namun, algoritma C4.5 lebih baik untuk menangani kasus ini dikarenakan tingkat akurasinya yang lebih tinggi dan minim error atau kesalahan, sehingga dapat memprediksi data dengan lebih tepat dan akurat.

5. Daftar Pustaka

- [1] Ardinata, M., 2020. Tanggung Jawab Negara Terhadap Jaminan Kesehatan Dalam Perspektif Hak Asasi Manusia. *Jurnal HAM*, 11(2), pp.319-332.
- [2] Inayati, S., Yuliana, Y. and Hanafiah, A., 2021. Prediksi Jumlah Peserta BPJS Penerima Bantuan Iuran (PBI) APBN menggunakan Metode Fuzzy Time Series Cheng. *Barekeng: Jurnal Ilmu Matematika dan Terapan*, 15(2), pp.373-384. DOI: <https://doi.org/10.30598/barekengvol15iss2pp373-384>
- [3] Nomor, P.P.R.I., 12. tahun 2013 tentang Jaminan Kesehatan. *Jakarta: Kementerian Hukum dan Hak Asasi Manusia*.
- [4] PERWALI Kota Depok No. 4 Tahun 2016 Tentang Petunjuk Teknis Pembiayaan Jaminan Kesehatan Masyarakat Miskin di Luar Kuota Penerima Bantuan Iuran Jaminan Kesehatan Dan Bantuan Sosial Tidak Terencana Bagi Orang Terlantar [JDIH BPK RI], n.d.)
- [5] Saputra, R.A., Wasiyanti, S. and Pribadi, D., 2021. Information Gain Pada Algoritma C4. 5 Untuk Klasifikasi Penerimaan Bantuan Pangan Non Tunai (BPNT). *Indonesian Journal of Business Intelligence (IJUBI)*, 4(1), pp.25-30. DOI: <http://dx.doi.org/10.21927/ijubi.v4i1.1757>
- [6] Sulihati, I., 2022. Penerapan Komparasi Algoritma C4. 5 dan Naive Bayes untuk Menentukan Hasil Seleksi Masuk Perguruan Tinggi. *JURNAL TECNOSCIENZA*, 6(2), pp.311-320.
- [7] Hakim, L.N., Sholihin, I., Rinaldi, A.R. and Morina, I.S., 2021. Penerapan Data Mining Algoritma Naive Bayes dalam Menentukan Program Keluarga Pra Sejahtera. *JURNAL DATA SCIENCE & INFORMATIKA*, 1(1), pp.21-25.
- [8] Ubaedi, I. and Djaksana, Y.M., 2022. Optimasi Algoritma C4. 5 Menggunakan Metode Forward Selection Dan Stratified Sampling Untuk Prediksi Kelayakan Kredit. *JSiI (Jurnal Sistem Informasi)*, 9(1), pp.17-26.
- [9] Bahtiar, A. and Silitonga, P.D., 2020. Penerapan Algoritma Decision Tree Untuk Memprediksi Penerima Bantuan Keluarga Harapan. *Jurnal ICT: Information Communication & Technology*, 19(1), pp.70-76. DOI: <https://doi.org/10.36054/jict-ikmi.v19i1.93>
- [10] Hasanah, M.A., Soim, S. and Handayani, A.S., 2021. Implementasi CRISP-DM Model Menggunakan Metode Decision Tree dengan Algoritma CART untuk Prediksi Curah Hujan Berpotensi Banjir. *Journal of Applied Informatics and Computing*, 5(2), pp.103-108.

- [11] Lidysari, W., Tambunan, H.S. and Qurniawan, H., 2022. Penerapan Data Mining Dalam Menentukan Kelayakan Penerima Bantuan Sosial Pemko Dengan Algoritma C4. 5 (Kasus Kantor Kelurahan Martoba). *Kesatria: Jurnal Penerapan Sistem Informasi (Komputer dan Manajemen)*, 3(1), pp.53-61.
- [12] Haryatmi, E. and Hervianti, S.P., 2021. Penerapan Algoritma Support Vector Machine Untuk Model Prediksi Kelulusan Mahasiswa Tepat Waktu. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 5(2), pp.386-392. DOI: <https://doi.org/10.29207/resti.v5i2.3007>
- [13] Algoritma Supervised Vs Unsupervised Learning, Cari Tahu Bedanya. 2022. DQLab. Available at: <https://dqlab.id/algoritma-supervised-vs-unsupervised-learning>
- [14] Tuntun, R., Kusriani, K. and Kusnawi, K., 2022. Analisis Perbandingan Kinerja Algoritma Klasifikasi dengan Menggunakan Metode K-Fold Cross Validation. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 6(4), pp.2111-2119.
- [15] Azhari, M., Situmorang, Z. and Rosnelly, R., 2021. Perbandingan Akurasi, Recall, dan Presisi Klasifikasi pada Algoritma C4. 5, Random Forest, SVM dan Naive Bayes. *Jurnal Media Informatika Budidarma*, 5(2), pp.640-651. DOI: <https://doi.org/10.30865/mib.v5i2.2937>
- [16] Damuri, A., Riyanto, U., Rusdianto, H. and Aminudin, M., 2021. Implementasi Data Mining dengan Algoritma Naive Bayes Untuk Klasifikasi Kelayakan Penerima Bantuan Sembako. *JURIKOM (Jurnal Riset Komputer)*, 8(6), pp.219-225. DOI: <https://doi.org/10.30865/jurikom.v8i6.3655>
- [17] Fitrihanah, D., Gunawan, W. and Sari, A.P., 2022. Studi Komparasi Algoritma Klasifikasi C5. 0, SVM dan Naive Bayes dengan Studi Kasus Prediksi Banjir. *Techno. Com*, 21(1), pp.1-11.