Journal of Information Technology Ampera
Vol. 7, No. 2, 2026 e-ISSN: **2774-2121**
https://journal-computing.org/index.php/journal-ita/index

Perbandingan Algoritma Klasifikasi *Decision Tree* dan *Naïve Bayes Classifier* untuk Prediksi Penyakit Diabetes

Andri[1], Iski meidiansyah[2], Muhammad bitragoya[3], Serly Oktarina[4]
[1]Sistem Informasi, Universitas Bina Darma, Palembang, Indonesia
[23,]Sistem Informasi ,Universitas Serelo, Lahat
[4]Ilmu Komputer, Universitas Sumatera Selatan
Email: [1]andri@binadarma.ac.id

Abstract

Diabetes merupakan salah satu penyakit yang mematikan yang harus menjadi perhatian bagi semua orang. Penyakit diabetes ini merupakan penyakit yang berlangsung sumur hidup serta dapat menimbulkan beberapa komplikasi jika penderita penyakit ini tidak bisa mengendalikannya. Sebagian besar penderita penyakit diabetes baru menyadari adanya penyakit ini setelah munculnya komplikasi serius seperti gangguan mata atau gangguan ginjal. Penelitian ini akan melakukan perbandingan algoritma klasifikasi *Decision Tree* dan *Naïve Bayes Classifier* dalam melakukan prediksi penyakit diabetes. Hasil akhir dari penelitian ini untuk melihat nilai akurasi dari masing-masing algoritma dalam melakukan prediksi berdasarkan model klasifikasi yang telah dihasil dari kedua algoritma tersebut. Dataset yang digunakan dalam melakukan uji coba kedua algoritma ini adalah *dataset* pasien penderita penyakit diabetes diperoleh dari repository dataset publik yang ada di situs resmi *Kaggle*. Berdasarkan hasil perhitungan nilai akurasi dari kedua model klasifikasi yang dibentuk dengan teknik sampling data *linear sampling* dan *shuffled* sampling, diperoleh nilai akurasi untuk model menggunakan algoritma *Decision Tree* sebesar 72.79% dan 71.87% dengan nilai *classification error* sebesar 27.21% dan 28.13%, sedangkan nilai akurasi menggunakan algoritma *Naïve Bayes Classifier* sebesar 76.05% dan 75.65% serta nilai *classification error* sebesar 23.95% dan 24.35%. Dari hasil ini dapat disimpulkan bahwa algoritma *naïve bayes* menghasilkan nilai akurasi yang sedikit lebih tinggi dibandingkan dengan algoritma *Decision Tree* dengan menggunakan *dataset* yang sama.

Keywords: Diabetes, data mining, *Decision Tree*, *Naïve Bayes Classifier*, klasifikasi

1. PENDAHULUAN

Mengutip dari situs resmi kementerian Kesehatan Republik Indonesia, International Diabetes Federation (IDF) menyatakan pada akhir tahun 2021

diabetes termasuk kedalam salah satu penyakit dengan pertumbuhan paling cepat, lebih dari setengah miliar manusia dari seluruh dunia hidup dengan diabetes atau tepatnya 537 juta orang dan jumlah ini akan diproyeksikan akan bertambah mencapai 643 juta orang pada tahun 2030 dan 783 juta pada tahun 2045[1].

Penyakit diabetes melitus adalah penyakit kronis yang terjadi karena peningkatan kadar gula dalam darah akibat tubuh tidak dapat menghasilkan atau menggunakan hormon insulin secara efektif. Penyakit diabetes melitus merupakan penyakit atau gangguan metabolisme kronis dengan multi etiologi yang ditandai dengan tingginya kadar gula darah disertai dengan gangguan metabolism karbohidrat, lipid dan protein sebagai akibat insufisiensi fungsi insulin[2].

Diabetes merupakan salah satu penyakit yang mematikan yang harus menjadi perhatian bagi semua orang. Penyakit diabetes ini merupakan penyakit yang berlangsung seumur hidup serta dapat menimbulkan beberapa komplikasi jika penderita penyakit ini tidak bisa mengendalikannya. Sebagian besar penderita penyakit diabetes baru menyadari adanya penyakit ini setelah munculnya komplikasi serius seperti gangguan mata atau gangguan ginjal.

Melihat bahayanya penyakit ini, maka sangat penting untuk mengetahui gejala diabetes sedini mungkin. Gejala umum yang sering dirasakan penderita diabetes adalah mudah lapar. Penderita diabetes akan merasakan lapar secara berlebihan sekalipun sudah makan dengan teratur. Hal ini kemungkinan disebabkan karena makanan yang dikonsumsi penderita diabetes sulit diubah menjadi energi akibat dari kekurangan hormon insulin.

Data mining merupakan teknik yang digunakan untuk menggali data dalam basis data berukuran besar. Data mining dapat kita gunakan untuk menemukan pengetahuan baru dari tumpukan data berukuran besar. Pola dari data histori yang tersimpan dalam sebuah basis data dapat ditemukan menggunakan teknik-teknik yang ada dalam data mining.

Secara sederhana data mining dapat didefinisikan sebagai kegiatan mengekstrak informasi atau pengetahuan penting dari suatu set data berukuran besar dengan menggunakan teknik tertentu untuk pengambilan keputusan dan memperkirakan perilaku dimasa yang akan datang[3]. Data mining disebut juga sebagai ilmu tentang pengumpulan data, pembersihan, pemrosesan, serta analisis bagaimana mendapatkan manfaat atau pengetahuan dari data[4].

Data mining dapat kita manfaatkan untuk melakukan prediksi penyakit dalam dunia kesehatan, seperti dalam penelitian ini, akan dilakukan penerapan teknik klasifikasi dalam data mining untuk melakukan prediksi penyakit diabetes yang diderita oleh pasien. Dengan memanfaatkan data historis pasien penderita diabetes terdahulu, data mining dapat melakukan prediksi dengan

cara membangun model prediksi berdasarkan atribut-atribut prediktor para penderita diabetes.

Dataset pasien penderita diabetes akan di training sehingga dapat dihasilkan sebuah model prediksi, berdasarkan model yang dihasilkan ini dapat diuji coba pada dataset pasien yang baru yang belum terdeteksi apakah pasien tersebut menderita diabetes atau tidak. Terdapat beberapa algoritma klasifikasi dalam data mining yang dapat digunakan untuk melakukan prediksi penyakit diabetes, diantaranya algoritma *Decision Tree* dan *Naïve Bayes Classifier*.

Decision Tree merupakan salah satu algoritma klasifikasi dalam data mining yang menggunakan seperangkat aturan untuk membuat keputusan dalam bentuk pohon (tree). Konsep dari algoritma ini yaitu mengubah data menjadi pohon keputusan. Algoritma ini memiliki kemampuan menguraikan pengambilan keputusan yang kompleks menjadi lebih sederhana, sehingga pengambilan keputusan akan lebih menginterpretasikan solusi dari permasalahan.

Naïve Bayes Classifier adalah algoritma klasifikasi dalam data mining yang memanfaatkan probabilitas dan statistik. Prinsip kerja dari algoritma ini adalah menemukan probabilitas atau kemungkinan suatu peristiwa akan terjadi dengan memberikan probabilitas peristiwa lain yang telah terjadi.

Fokus dari penelitian ini yaitu melakukan perbandingan algoritma klasifikasi *Decision Tree* dan *Naïve Bayes Classifier* dalam melakukan prediksi penyakit diabetes dengan menggunakan dataset yang sama. Hasil akhir dari penelitian ini untuk melihat nilai akurasi dari masing-masing algoritma dalam melakukan prediksi berdasarkan model klasifikasi yang telah terbentuk. Dataset yang digunakan dalam melakukan uji coba kedua algoritma ini diambil dari repository dataset publik yang ada di situs resmi Kaggle.

Terdapat beberapa penelitian terdahulu yang telah menerapkan algoritma *Decision Tree* dan naïve bayes dalam melakukan prediksi penyakit jantung. Diantaranya penelitian dengan judul Studi Komparasi Algoritma ID3 dan Algoritma Naïve Bayes untuk Klasifikasi Penyakit Diabetes Mellitus[5]. Dalam penelitian ini dibahas perbandingan algoritma ID3 dan Naïve Bayes untuk melakukan klasifikasi penyakit diabetes mellitus. Hasil klasifikasi dari penelitian ini dengan menggunakan Confusion Matrix dan kurva ROC diperoleh nilai akurasi untuk algoritam ID3 sebesar 74% dan nilai akurasi algoritma Naïve Bayes sebesar 76%. Dari penelitian ini dapat diketahui bahwa algoritma Naïve Bayes memiliki tingkat akurasi prediksi lebih tinggi.

Selanjutnya penelitian dengan judul Komparasi Akurasi Metode Correlated *Naïve Bayes Classifier* dan *Naïve Bayes Classifier* untuk Diagnosis Penyakit Diabetes[6]. Penelitian ini melakukan perbandingan dua buah algoritma

klasifikasi Correlated Naïve Bayes dan *Naïve Bayes Classifier*. Dari hasil perbandingan kedua algoritma ini diketahui nilai akurasi tertinggi diperoleh dengan menggunakan algoritma Correlated Naïve Bayes sebesar 67.15% sedangkan nilai akurasi algoritmat *Naïve Bayes Classifier* diperoleh nilai akurasi sebesar 67.15%.

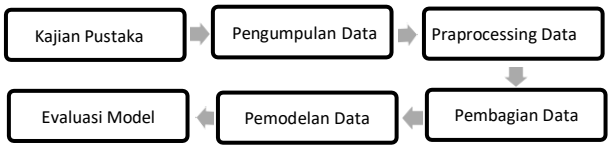
Dilihat dari penelitian terdahulu yang telah dilakukan, terdapat beberapa perbedaan utama penelitian ini dari penelitian sebelumnya yang membahas tentang perbandingan algoritma klasifikasi, yaitu:

a. Penelitian ini terdapat tahapan seleksi fitur dengan menggunakan teknik pengukuran bobot gain rasio dari masing-masing fitur. Penggunaan teknik ini dapat mereduksi dimensi fitur dengan cara mengukur reduksi entropy sebelum dan sesudah pemisahan. Manfaat dari seleksi fitur akan mempercepat proses training data yang dilakukan, selain itu seleksi fitur juga digunakan untuk mengukur seberapa relevan atau berpengaruh atribut atau fitur terhadap hasil pengukuran.
b. Penelitian ini mengukur nilai recall, precision dan f-measure untuk melihat tingkat performan dari model klasifikasi yang dihasilkan.

Pengukuran performan algoritma menggunakan metode k-fold validation yang menerapkan dua teknik sampling data, linear sampling dan shuffled sampling.

2. METODE

Ada beberapa langkah yang akan dilalui dalam penelitian ini, yang pertama adalah melakukan kajian pustaka, yang kedua melakukan pengumpulan data, ketiga adalah pra-pemrosesan (preprocessing) data, keempat pembagian data, kelima proses pemodelan data dan keenam adalah melakukan evaluasi model yang dihasilkan. Langkah-langka dalam penelitian ini dapat dilihat seperti digambarkan pada gambar 1.



Gambar 1: Tahapan Penelitian

2.1. Kajian Pustaka

Tahapan kajian pustaka dilakukan untuk mempelajari berbagai referensi serta mengumpulkan informasi yang terkait dalam penelitian. Pemahaman tentang konsep data mining diperlukan untuk mempermudah dalam menyelesaikan permasalahan yang telah didefinisikan. Dalam tahapan ini akan dipelajari literatur yang berhubungan dengan algoritma klasifikasi dalam data mining yang diantaranya algoritma *Decision Tree* dan *Naïve Bayes Classifier*.

2.2. Pengumpulan Data

Tahapan kedua dari penelitian adalah mengumpulkan data. Dataset yang digunakan dalam penelitian ini merupakan dataset publik yaitu dataset penyakit diabetes yang diambil dari website resmi kaggle yang dapat diakses melalui alamat web
https://www.kaggle.com/datasets/akshaydattatraykhare/diabetes-dataset.
Dataset ini merupakan data original dari National Institute of Diabetes and Digestive and Kidney Diseases. Tujuan dari dataset ini adalah untuk memprediksi secara diagnostic apakah seorang pasien menderita diabetes berdasarkan pengukuran diagnostic tertentu.

2.3. Praprocessing Data

Sebelum data dimodelkan, tahapan yang perlu dilakukan adalah tahapan pra-processing data. Tahapan ini bertujuan untuk melakukan proses pembersihan data (*cleaning*) yang bertujuan untuk mengetahui apakah data yang akan digunakan sudah valid atau belum. Pembersihan data merupakan suatu teknik yang digunakan dalam menangani data kosong atau data dengan nilai yang tidak lengkap. Pembersihan data biasanya juga dilakukan untuk melakukan standarisasi format data yang digunakan sebagai dataset.
Jika proses pembersihan data telah dilakukan maka langkah berikutnya adalah melakukan proses pemilihan fitur (*feature selection*). Pemilihan fitur bertujuan untuk memilih atribut-atribut atau fitur mana yang paling dominan yang akan digunakan untuk proses pemodelan data.
Tahapan seleksi fitur dilakukan pengukuran dari setiap fitur guna untuk mengetahui fitur yang dominan yang akan dipakai dalam pemodelan data.

Pada penelitian ini, pemilihan fitur dilakukan dengan cara menghitung nilai bobot (*weight*) berdasarkan nilai information gain rasio dari setiap fitur. Semakin tinggi bobot gain rasio dari suatu atribut, maka semakin relevan atribut tersebut untuk dipertimbangkan.

2.4. Pembagian Data

Setelah proses praprocessing data dilakukan, langkah selanjutnya adalah pembagian data. Pembagian data ini bertujuan untuk membagi dataset menjadi data training dan data testing. Data training nantinya akan digunakan untuk melakukan training dataset dengan menggunakan algoritma klasifikasi *Decision Tree* dan *Naïve Bayes Classifier*. Dalam melakukan pembagian dataset menjadi data training dan testing digunakan dua tipe teknik sampling, yaitu teknik linear sampling dan *shuffled sampling*. *Linear sampling* dilakukan dengan cara mengambil sampel dengan cara membagi dataset menjadi beberapa partisi tanpa mengubah urutannya. Sedangkan taknik *shuffled* sampling dilakukan pengambilan sampel secara acak dari dataset.

2.5. Pemodelan Data

1. Decision Tree

Decision Tree merupakan salah satu model dalam data mining untuk melakukan klasifikasi prediksi dalam bentuk pohon (*tree*). *Decision Tree* digunakan untuk pengenalan pola dan termasuk ke dalam pengenalan pola secara statistik. *Decision Tree* menghasilkan pohon yang dibangun dalam suatu metode rekursif *top-down divide and conquer*.

Algoritma *Decision Tree* yang digunakan dalam penelitian ini adalah algoritma C4.5, algoritma ini merupakan pengembangan dari algoritma ID3 (*Iterative Dichotomiser*) yang mengadopsi *greedy* atau non backtracking dimana tree dibangun dengan teknik *top-down* dan diulang (*recursive*) serta dibagi, kemudian diselesaikan[7]. Tree yang terbentuk dari algoritma C4.5 ini memiliki kelebihan dalam menangani overfiting dengan cara melakukan pruning.

Dalam membangun tree, algoritma C4.5 pertama akan menentukan root yang didapat dari nilai gain yang paling tinggi diantara atribut-atribut yang ada didalam dataset. Sebelum menentukan nilai gain terlebih dahulu algoritma ini akan mencari nilai *entropy* dari masing-masing atribut. Rumus perhitungan nilai gain dan entropy dapat dilihat pada gambar 2.

$$\text{Gain(S,A)} = Entropy(S) - \sum_{i=1}^{n} \frac{|Si|}{S} * Entropy(Si)$$

$$\text{Entropy(S)} = \sum_{i=1}^{n} - pi * \log_2 pi$$

Gambar 2. Rumus Gain dan Entropy

Keterangan:

S : Himpunan Kasus
A : Atribut
n : Jumlah partisi S
|Si| : Jumlah kasus pada partisi ke i
|S| : Jumlah kasus dalam S
|pi| : Proporsi dari Si terhadap S

2. *Naïve Bayes Classifier*

Naïve Bayes Classifier adalah metode probabilistik yang digunakan dalam berbagai permasalahan klasifikasi. Dasar dari algoritma naïve bayes adalah teorema naïve bayes. Dalam teorema naïve bayes kita dapat mengetahui sebuah probabilitas dengan cara menghitung probabilitas lain yang berkaitan. *Naïve Bayes Classifier* merupakan algoritma klasifikasi yang berakar pada teorema Bayes. Teorema Bayes merupakan teknik prediksi berdasarkan kemungkinan[8]. Rumus umum dari prediksi bayes adalah:

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

Keterangan:

X : Data testing yang kelasnya belum diketahui
H : Hipotesis data X yang merupakan suatu kelas yang lebih spesifik
P(H|X) : Probabilitas hipotesis berdasar kondisi (*posterriori probability*)
P(X|H) : Probabilitas hipotesis X berdasarkan kondisi H (*likelihood*)
P(H) : Probabilitas hipotesis H (*prior probability*)
P(X) : Probabilitas hipotesis X (*predictor prior probability*)

2.6. Evaluasi Model

Evaluasi model merupakan tahapan yang dilakukan untuk mengetahui seberapa baik performa dari model yang dihasilkan dari uji coba data training dan data testing. Untuk melihat performa dari algoritma klasifikasi dalam data mining dapat dilihat dari nilai akurasi yang dihasilkan oleh model yang digunakan.

3. HASIL DAN PEMBAHASAN

Penelitian ini melakukan perbandingan algoritma *Decision Tree* dan *Naïve bayes* untuk melakukan prediksi pasien penderita penyakit diabetes. Dataset awal terdiri dari 768 record dan terdapat 8 atribut regular serta 1 atribut kelas. Atribut-atribut yang ada dalam dataset diabetes sebelum dilakukan pemilihan fitur dapat dilihat dalam tabel 1.

Tabel 1. Atribut/fitur Dataset Diabetes

Fitur/atribut	Keterangan
BloodPressure	Atribut Regular
Age	Atribut Regular
Insulin	Atribut Regular
DiabetesPedigreeFunction	Atribut Regular
SkinThickness	Atribut Regular
BMI	Atribut Regular
Pregnancies	Atribut Regular
Glucose	Atribut Regular
Outcome	Kelas/Label

Pemilihan fitur atau atribut dalam penelitian ini dilakukan dengan menghitung nilai bobot dari information gain ratio dari setiap atribut. Berdasarkan hasil perhitungan information gain ratio setiap fitur dapat ditentukan jumlah atribut yang digunakan dalam penelitian sebanyak 6 atribut regular dan 1 atribut spesial sebagai kelas atau label. Dataset awal memiliki jumlah fitur sebanyak 9 fitur serta jumlah record atau baris sebanyak 768 baris. Setelah dilakukan proses pemilihan fitur dengan menghitung nilai bobot berdasarkan informasi gain ratio dan setelah ditentukan threshold sebesar 0.1, maka fitur yang memiliki bobot information gain ratio dibawah 0.1 tidak digunakan dalam proses pemodelan data. Berdasarkan hasil penghitungan menggunakan tools Rapidminer bobot information gain ratio untuk masing-masing fitur diperlihatkan pada tabel 2.

Tabel 2. Nilai bobot fitur berdasarkan gain ratio

Fitur	Bobot Information Gain Ratio
BloodPressure	0.056
Age	0.073
Insulin	0.138
DiabetesPedigreeFunction	0.138
SkinThickness	0.152
BMI	0.152
Pregnancies	0.169
Glucose	0.196

Pengukuran performa kedua algoritma menggunakan teknik *k-fold cross validation*. *Cross validation* merupakan metode yang bertujuan untuk memperoleh hasil akurasi yang maksimal dari dari suatu model. Dalam *k-fold cross validation* percobaan dilakukan sebanyak k kali untuk suatu model dengan parameter yang sama[3]. Dalam penelitian ini nilai k didefinisikan sebesar 10, ini berarti perulangan pembagian data training dan data testing sebanyak 10 kali dengan tipe sampling data training dan testing menggunakan tipe linear sampling dan shuffled sampling.

Berdasarkan hasil analisis pembentukan model menggunakan *tools RapidMiner*, diperoleh hasil akurasi untuk masing-masing algoritma tidak terlalu berbeda signifikan. Setelah dilakukan pemodelan menggunakan algoritma klasifikasi *Decision Tree* dan *Naïve Bayes Classifier* dengan metode linear sampling data, diperoleh nilai akurasi untuk algoritma *Decision Tree* sebesar 72.79% dengan tingkat classification error sebesar 27.21%, sedangkan algoritma *Naïve Bayes Classifier* menghasilkan nilai akurasi sebesar 76.05% dengan tingkat *classification error* sebesar 23.95%. Hasil perbandingan nilai akurasi dan *classification error* untuk kedua algoritma ini diperlihatkan pada tabel 3.

Tabel 3. Perbandingan Nilai Akurasi dan *Classification Error* (Liner Sampling)

Metode	Accuracy	Classification Error

Decision Tree	72.79%	27.21%
Naïve Bayes Classifier	76.05%	23.95%

Untuk mengetahui hasil nilai akurasi yang diperoleh dapat menggunakan tabel dalam bentuk matrik atau disebut juga dengan *confusion matrix*. *Confusion matrix* memberikan informasi perbandingan hasil klasifikasi yang dihasilkan oleh model dengan hasil klasifikasi data sebenarnya. Rumus untuk menghitung akurasi menggunakan confusion matrix ditunjukan oleh gambar 3.

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

Gambar 3. Rumus menghitung akurasi

Nilai akurasi sebesar 72.79% yang diperoleh dari pengujian dataset menggunakan algoritma *Decision Tree* diperoleh dari tabel confusion matrix yang ditunjukkan pada tabel 4.

Tabel 4. Confusion Matrix *Decision Tree*

	True 1	**True 0**
Pred. 1	100	41
Pred. 0	168	459

Untuk nilai akurasi 76.05% yang diperoleh berdasarkan hasil pengujian data training dan data testing menggunakan algoritma *Naïve Bayes Classifier* diperoleh dari tabel confusion matrix yang digambarkan pada tabel 5.

Tabel 5. Confusion Matrix Naïve Bayes

	True 1	True 0
Pred. 1	154	70
Pred. 0	114	430

Dari tabel confusion matrix yang telah dihasilkan, dapat dihitung nilai recall, precision dan f-measure. Hasil lengkap perbandingan antara nilai-nilai tersebut dengan menggunakan tipe sampling linear dan shuffled untuk kedua algoritma dimana nilai k-fold cross validation ditetapkan k =10 dapat di lihat pada tabel 6 dan tabel 7.

Tabel 6. Perbandingan nilai accuracy, recal, precision dan f-measure (Linear Sampling)

Algoritma	Accuracy	Recall	Precision	F-Measure
DTC4.5	72.79%	91.69%	73.04%	81.15%
Naïve Bayes	76.05%	85.94%	78.84%	82.10%

Tabel 7. Perbandingan nilai accuracy, recall, precision dan f-measure (Shufled Sampling)

Algoritma	Accuracy	Recall	Precision	F-Measure
DTC4.5	71.87%	92.62%	72.07%	80.93%
Naïve Bayes	75.65%	85.47%	78.91%	81.98%

4. KESIMPULAN

Berdasarkan hasil perbandingan antara kedua algoritma klasifikasi untuk prediksi pasien penderita penyakit diabetes, nilai akurasi untuk algoritma *Naïve Bayes Classifier* sedikit lebih tinggi dibandingkan klasifikasi *Decision Tree* yang menggunakan algoritma C4.5. Dengan menggunakan teknik sampling dataset linear *sampling* dan *shuffled* sampling, algoritma *Naïve Bayes Classifier* memiliki akurasi sebesar 76.05% dan 75.65% sedangkan algoritma C4.5 menghasilkan nilai akurasi sebesar 72.79% dan 71.87%. Berbeda dengan nilai *Recall*, Algoritma *Decision Tree* memiliki nilai *Recall* sedikit lebih tinggi dari algoritma *Naïve Bayes Classifier*. Sedangkan nilai *Precision* dan *F-Measure* algoritma *Naïve Bayes Classifier* sedikit lebih tinggi dari *Decision Tree*.

DAFTAR PUSTAKA

[1] M. R. Saraswati, "Diabetes Melitus adalah Masalah Kita," *Kementrian Kesehatan RI*, 2022. https://yankes.kemkes.go.id/view_artikel/1131/diabetes-melitus-adalah-masalah-kita (accessed Feb. 24, 2023).

[2] WHO, "Diabetes," 2022. https://www.who.int/health-topics/diabetes#tab=tab_1 (accessed Feb. 24, 2023).

[3] A. Santosa, B., & Umam, *Data Mining dan Big Data Analytics*. Yogyakarta: Penebar Media Pustaka, 2018.

[4] C. C. Aggarwal, *Data Mining*. New: Springer International Publishing Switzerland, 2015.

[5] N. Nunu and A. Abijar, "Studi Komparasi Algoritma ID3 dan Algoritma NAIVE BAYES untuk Klasifikasi Penyakit Diabetes Mellitus," pp. 18–23, 2015.

[6] Hairani, G. S. Nugraha, M. N. Abdillah, and M. Innuddin, "Komparasi Akurasi Metode Correlated Naive Bayes Classifier Dan Naive Bayes Classifier Untuk Diagnosis Penyakit Diabetes," *J. Nas. Inform. dan Teknol. Jar.*, pp. 6–11, 2018.

[7] J. Han and M. Kamber, *Data Mining: Concept and Techniques Second Edition*. Morgan Kaufmann Publishers, 2006.

[8] M. Guntur, J. Santony, and Yuhandri, "Prediksi Harga Emas dengan Menggunakan Metode Naïve Bayes," *RESTI*, vol. 2, no. 1, pp. 354–360, 2018.