Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 5, No. 1, November 2023, Page 165-174
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v5i1.4614

Perbandingan Algoritma Klasifikasi Data Mining Pada Prediksi Penyakit Diabetes

Yunan Fauzi Wijaya[1], Agung Triayudi[2,*]

[1] Fakultas Teknologi Komunikasi dan Informatika, Informatika, Universitas Nasional, Jakarta, Indonesia
[2] Fakultas Teknologi Komunikasi dan Informatika, Teknologi Informasi, Universitas Nasional, Jakarta, Indonesia
Email: [1]yunan.fw@civitas.unas.ac.id, [2,*]agungtriayudi@civitas.unas.ac.id
Email Penulis Korespondensi: agungtriayudi@civitas.unas.ac.id
Submitted: 23/11/2023; Accepted: 30/11/2023; Published: 30/11/2023

Abstrak−Diabetes merupakan sebuah penyakit kronis yang menyerang manusia. Salah satu penyebab dari pada terjadinya penyakit diabetes pada manusia karena asupan gula terlalu tinggi yang tidak dapat diseimbangi oleh tubuh terkait dengan penyerapan ataupun aktivitas yang dilakukan. Penyakit diabetes sering kali dianggap sebagai penyakit biasa dikalangan masyarakat, tetapi dampak yang terjadi diakibatkan oleh penyakit ini sangatlah merugikan bagi manusia. Berdasarkan hal tersebut maka kiranya perlu diketahui bagi setiap orang menderita penyakit diabetes atau tidak. Maka dari itu, permasalahan tersebut harus diselesaikan dengan tepat, dimana perlu dilakukan prediksi terhadap sesorang terkena penyakit diabetes atau tidak. Proses prediksi yang dilakukan untuk menentukan sesorang terkena penyakit diabtes atau tidak dengan mengetahui pola ataupun gejala – gejala kemungkinan yang menyebabkan seseorang menderita penyakit diabetes. Pada penelitian ini proses pembentukan pola berdasarkan dengan data – data yang tersimpan dimasa lampau yang terkumpul pada dataset. Dataset merupakan sebuah kumpulan – kumpulan data dimasa lampau yang terjadi secara fakta dan kemudian dikumpulkan dalam jangka waktu tertentu secara besar. Data mining merupakan sebuah cara yang dipergunakan untuk melakukan pengolahan data berdasarkan dengan kumpulan – kumpulan data dimasa lampau baik pada dataset ataupun yang lainnya. Pada data mining proses pengolahan data dilakukan dengan berbagai macam teknik, salah satu yang menjadi teknik penyelesaian pada data mining adalah klasifikasi. Pada penelitian ini akan digunakan algoritma Naïve Bayes, algortima K-Nearest Neighbor (K-NN) dan algoritma C4.5. Pada proses klasifikasi data mining terdapat 3 (tiga) algoritma yang digunakan yaitu Naïve Bayes, K-Nearest Neighbor dan C4.5 Dari hasil pengujian yang telah dilakukan didapatkan hasil kinerja akurasi untuk algoritma Naïve Bayes sebesar 75%, akurasi untuk algoirtma K-Nearest Neighbor sebsar 80,60% dan algoritma C4.5 sebsar 91,80%. Dalam hal ini menandakan bahwasannya algoritma C4.5 memiliki kinerja lebih baik dibandingan dengan algroritma lainnya. Maka dari itu hasil pola yang dihasil dari algoritma C4.5 dipergunakan untuk melakukan prediksi terhadap penyakit diabtes tersebut.

Kata Kunci: Perbandingan; Klasifikasi; Data Mining; Prediksi; Diabetes

Abstract−Diabetes is a chronic disease that attacks humans. One of the causes of diabetes in humans is that sugar intake is too high which the body cannot balance due to absorption or activities carried out. Diabetes is often considered a common disease among people, but the impacts caused by this disease are very detrimental to humans. Based on this, it is necessary for everyone to know whether they suffer from diabetes or not. Therefore, this problem must be resolved appropriately, where it is necessary to predict whether someone will have diabetes or not. The prediction process is carried out to determine whether someone has diabetes or not by knowing the patterns or possible symptoms that cause someone to suffer from diabetes. In this research, the pattern formation process is based on data stored in the past collected in a dataset. A dataset is a collection of past data that occurred in fact and was then collected over a certain period of time on a large scale. Data mining is a method used to process data based on collections of past data, whether in datasets or others. In data mining, the data processing process is carried out using various techniques, one of which is the solution technique in data mining is classification. In this research, the Naïve Bayes algorithm, the K-Nearest Neighbor (K-NN) algorithm and the C4.5 algorithm will be used. In the data mining classification process, there are 3 (three) algorithms used, namely Naïve Bayes, K-Nearest Neighbor and C4.5. From the results of the tests that have been carried out, the accuracy performance results for the Naïve Bayes algorithm are 75%, accuracy for the K-Nearest Neighbor algorithm by 80.60% and the C4.5 algorithm by 91.80%. In this case, it indicates that the C4.5 algorithm has better performance compared to other algorithms. Therefore, the pattern results produced by the C4.5 algorithm are used to make predictions about diabetes.

Keywords: Comparison; Classification; Data Mining; Prediction; Diabetes

1. PENDAHULUAN

Diabetes merupakan sebuah penyakit kronis yang menyerang manusia. Penyakit diabetes disebabkan oleh tingginya kadar gula dalam tubuh manusia. Penyakit diabetes menyerang dari pada fungsi metabolisme pada tubuh dimana tubuh tidak dapat mencerna ataupun menggunakan kadar gulu yang tinggi berada pada tubuh manusia. Penyakit diabetes sudah tergolong penyakit kronis yang berbahaya dikarenakan mengakibatkan dampak yang sangat fatal bagi manusia terutama jika sudah terjadi komplikasi penyakit[1], [2].

Salah satu penyebab dari pada terjadinya penyakit diabetes pada manusia karena asupan gula terlalu tinggi yang tidak dapat diseimbangi oleh tubuh terkait dengan penyerapan ataupun aktivitas yang dilakukan. Hal tersebut lah yang menyebabkan penumpukan terhadap gula darah di tubuh yang menyebabkan terjadinya penyakit diabtes. Pada umumnya penyakit diabetes terbagi atas beberapa jenis mulai dari yang dapat diobati hingga yang dapat menyebabkan kompilkasi[3], [4].

Penyakit diabetes sering kali dianggap sebagai penyakit biasa dikalangan masyarakat, tetapi dampak yang terjadi diakibatkan oleh penyakit ini sangatlah merugikan bagi manusia. Dimulai dari terjadi jenis penyakit –

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 5, No. 1, November 2023, Page 165-174
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v5i1.4614

penyakit lainnya hingga terjadi kematian dikarenakan kompilkasi yang terjadi dari beberapa penyakit terjadi secara bersamaan diderita oleh manusia[5], [6].

Berdasarkan hal tersebut maka kiranya perlu diketahui bagi setiap orang menderita penyakit diabetes atau tidak. Dengan mengetahui seseorang menderita penyakit diabetes atau tidak agar kiranya dapat diberikan penangan awal sedari dini agar penyakit diabtes tersebut tidak menyebar ataupun bahkan bertambah parah di tubuh manusia yang menyebabkan dampak begitu fatal.

Maka dari itu, permasalahan tersebut harus diselesaikan dengan tepat, dimana perlu dilakukan prediksi terhadap sesorang terkena penyakit diabetes atau tidak. Prediksi yang dilakukan untuk memperikiran bagi sesorang menderita penyakit diabetes ataupun seseorang tersebut memiliki kemungkinan untuk terjangkit penyakit diabetes tersebut.

Proses prediksi yang dilakukan untuk menentukan sesorang terkena penyakit diabtes atau tidak dengan mengetahui pola ataupun gejala – gejala kemungkinan yang menyebabkan seseorang menderita penyakit diabetes. Dengan mengetahui bagaimana pola – pola atau atribut seseorang yang terkena penyakit diabtes dapat kiranya jika terdapat seseorang yang menderita penyakit diabetes maka dapat ditangani dengan segera[7], [8].

Pada penelitian ini proses pembentukan pola berdasarkan dengan data – data yang tersimpan dimasa lampau yang terkumpul pada dataset. Dalam hal ini, penelitian ini menggunakan dataset diabetes untuk membentuk pola – pola yang dapat dipergunakan untuk melakukan prediksi terhadap penyakit diabetes tersebut. Dengan dataset tersebut nantinya didapatkan informasi yang dapat dipergunakan dalam proses penelitian.

Dataset merupakan sebuah kumpulan – kumpulan data dimasa lampau yang terjadi secara fakta dan kemudian dikumpulkan dalam jangka waktu tertentu secara besar. Pada dataset data yang tersimpan dapat diproses kembali sesuai dengan kebutuhan terhadap penggunaan data pada dataset tersebut. Penggunaan data pada dataset disesuaikan kembali dengan hasil proses seperti yang diinginkan oleh pemilik data. Maka dari itu perlu kiranya dibutuhkan sebuah cara ataupun teknik khusus yang dipergunakan untuk membantu dalam melakukan proses pengolahan data pada dataset. Cara tersebut biasa digunakan dengan data mining[9], [10].

Data mining merupakan sebuah cara yang dipergunakan untuk melakukan pengolahan data berdasarkan dengan kumpulan – kumpulan data dimasa lampau baik pada dataset ataupun yang lainnya. Proses yang dilakukan pada data mining bertujuan untuk menghasilkan sebuah informasi – informasi baru yang berharga dan kemudian dapat dipergunakan kembali bagi pemilik data ataupun pemroses data. Informasi yang didapatkan pada data minng berdasarkan dengan pola hubungan yang terbentuk dari hasil pemrosesan data mining[11], [12].

Pada data mining proses pengolahan data dilakukan dengan berbagai macam teknik, salah satu yang menjadi teknik penyelesaian pada data mining adalah klasifikasi. Klasifikasi merupakan proses pada data mining yang berdasarkan dengan pengelompokan data berdasarkan dengan kriteria – kriteria tertentu. Pada data mining, klasifikasi bertujuan untuk mengelompokan data terhadap kelas – kelas tertentu sesuai dengan kriteria – kriteria yang telah ditentukan. Kriteria pada klasifikasi sangat berpengaruh terhadap proses pengelompokan atau klasifikasi data. Dari kriteria tersebut nantinya pada klasifikasi akan terbentuk hubungan pola – pola yang dapat digunakan untuk proses penyelesaian masalah[13]–[15].

Pada klasifikasi data mining terdapat berbagai macam algoritma yang digunakan untuk menyelesaikan masalah. Pada penelitian ini akan digunakan algoritma Naïve Bayes, algortima K-Nearest Neighbor (K-NN) dan algoritma C4.5. Ketiga algoritma tersebut merupakan algoritma – algoritma yang sering digunakan untuk proses penyelesaian pada klasifikasi data mining.

Algoritma Naïve Bayes merupakan bagian dari pada klasifikasi pada data mining yang mengelompokan data berdasarkan dengan nilai probabilitas dari setiap kemungkinan data berdasarkan dengan tujuan kelasnya. Nilai probabilitas yang dihasilkan pada Naïve Bayes berdasarkan dengan pengembangan dari pada teorema bayes yang melakukan perhitungan terhadap beberapa kemungkinan yang terjadi dari sebuah peristiwa[16]–[18].

Sebagai dasar dalam proses penelitian maka diperlukan beberapa penelitian terdahulu seperti yang dilakukan oleh Nur Syifa Fauzia dan Raditya Danar Dana Program pada tahun 2023 dengan judul penelitian "Implementasi Algoritma Naive bayes dalam Klasifikasi Status Kesejahteraan Masyarakat Desa Gunungsari" dimana hasil penelitian yang didapatkan bahwasannya algoritma Naïve Bayes memiliki kinerja sebesar 93,69% dalam melakukan klasifikasi status kesejahteraan masyrakat[19].

Penelitian lainnya yang dilakukan oleh Ika Nurjanah, dkk pada tahun 2023 juga dengan judul penelitian yang dilakukan "Penggunaan Algoritma Naïve Bayes Untuk Menentukan Pemberian Kredit Pada Koperasi Desa" dari hasil penelitian yang dilakukan mengatkan bahwasannya algoritma tersebut dapat dipergunakan untuk memberikan kelayakan terhadap nasabah dalam pemberian kredit[20].

Penelitian terakhir pada Naïve Bayes sebagai penelitian terdahulu seperti yang dilakukan oleh Qurotul A'yuniyah, dkk pada tahun 2023 dengan judul penelitian "Analisa Algoritma Naïve Bayes Classifier (NBC) Untuk Prediksi Penjualan Alat Kesehatan" dimana pada proses yang dilakukan algoritma Naïve Bayes memiliki kinerja sebesar 95% dengan terdapat 2 kelas yang dihasilkan[21].

Algoritma lainnya yang merupakan bagian dari pada klasifikasi data mining adalah algoritma K-Nearest Neighbor (K-NN). Algoritma K-Nearest Neighbor (K-NN) digunakan sebagai klasifikasi pada data mining berdasarkan dengan nilai kedekatan jarak dari setiap objek. Proses pertama yang dilakukan dari algoritma K-Nearest Neighbor (K-NN) yaitu penentuan nilai K awal sebagai dasar pengambilan keputusan nantinya.

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 5, No. 1, November 2023, Page 165-174
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v5i1.4614

Perhitungan nilai jarak dari pada algoritma K-Nearest Neighbor (K-NN) berdasarkan dengan nilai euclidean distance dari objek baru dengan objek yang sebelumnya[22]–[24].

Terdapat juga penelitian sebagai dasar dalam pelaksanaan penelitian dengan algoritma K-Nearest Neighbor (K-NN) seperti yang dilakukan oleh Ihzan Sayid Muallif, dkk pada tahun 2023 "Penerapan Data Mining untuk Prediksi Pergerakan Harga Saham Menggunakan Algoritma K-Nearest Neighbor" dari hasil penelitian yang dilakukan bahwasannya algoritma K-Nearest Neighbor (K-NN) memiliki kinerja sebesar 62,54% dengan precision sebesar 64,14% dan recall sebesar 92,08%[25].

Terdapat juga penelitian lainnya yang dilakukan oleh Rahmadini, dkk pada tahun 2023 dengan judul penelitian "Penerapan Data Mining Untuk Memprediksi Harga Bahan Pangan Di Indonesia Menggunakan Algoritma K-Nearest Neighbor" dimana didapatkan hasil penelitian bahwasannya dengan pengujian terbaik pada nilai K=2 dengan MAE dan RMSE untuk data training 52,77 dan 96,40 dan untuk data testing 55,55 dan 81,64[26].

Penelitian lainnya seperti yang dilakukan oleh Rino Bahtiar pada tahun 2023 dengan judul "Implementasi Data Mining Untuk Prediksi Penjualan Kusen Terlaris Menggunakan Metode K-Nearest Neighbor" dimana hasil penelitian yang didapatkan bahwasannya tingkat akurasi yang didapatkan dari proses penelitian sebesar 88,89% pada data penjualan serta 80% untuk data penjualan[27].

Algoritma terakhir yang digunakan pada penelitian ini untuk melakukan klasfikasi adalah algoritma C4.5. Algoritma C4.5 merupakan bagian dari klasifikasi data mining berdasarkan dengan pembentukan pohon keputusan. Pohon keputusan yang dibentuk dari algoritma C4.5 nantinya menghasilkan sebuah rule baru, rule tersebut yang dipergunakan sebagai dasar proses pengambilan keputusan. Pembentukan pohon keputusan pada algoritma C4.5 berdasarkan dengan perhitungan nilai gain dan entropy. Atribut yang memiliki nilai gain tertinggi nantinya akan digunakan sebagai akar utama, hingga nanti tidak lagi terdapat percabangan terhadap proses pengambilan keputusan[28]–[30].

Terdapat penelitian terdahulu pada algoritma C4.5 seperti yang dilakukan oleh Fery Pirmansyah dan Tri Wahyudi pada tahun 2023 dengan judul penelitian "Implementasi Data Mining Menggunakan Algoritma C4.5 Untuk Prediksi Evaluasi Anggota Satuan Pengamanan Studi Kasus PT. YIMM Pulogadung" dimana hasil yang didapatkan dari penelitian bahwa tingkat akurasi yang didapatkan dari proses pengujian sebesar 99,84%[31].

Penelitian lainnya yang juga dilakukan pada tahun 2023 oleh Sandy Mulyanda, dkk dengan judul penelitian "Analisis Data Mining Menggunakan Algoritma C4.5 Untuk Prediksi Harga Pasar Mobil Bekas" didapatkan hasil penelitian bahwasannya algoritma C4.5 dapat dipergunakan untuk melakukan prediksi terhadap harga pasar mobil dengan tingkat akurasi yang didapatkan sebesar 99%[32].

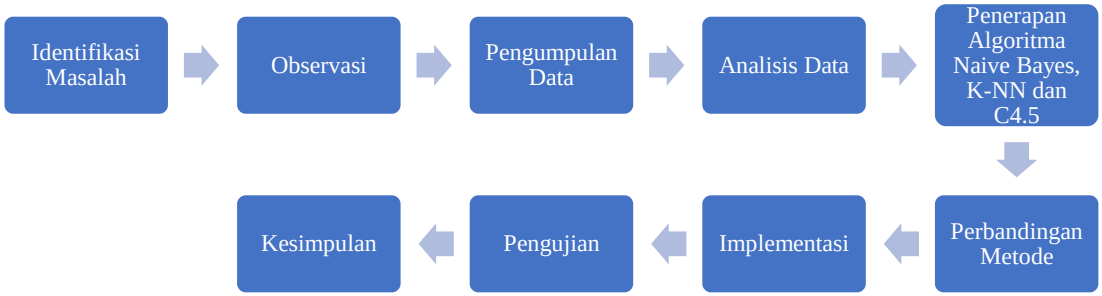
Penelitian terakhir yang digunakan sebagai referensi penelitian yang dilakukan oleh Tri Widiastuti, dkk pada tahun 2023 dengan judul penelitian "Evaluasi Tingkat Kepuasan Mahasiswa Terhadap Pelayanan Akademik Menggunakan Metode Klasifikasi Algoritma C4.5" dimana hasil yang didapatkan dari penelitian nilai akurasi yang didapatkan dari proses pengujian sebesar 90%[33].

Tujuan penggunaan dari ketiga algoritma klasifikasi data mining tersebut untuk dilakukan proses perbandingan kinerja algoritma. Perbandingan algoritma bertujuan untuk mendapatkan algoritma yang memiliki kinerja paling baik berdasarkan dengan tingkat akurasi yang dihasilkan. Dari hasil algoritma yang memiliki kinerja yang paling baik nantinya akan digunakan sebagai dasar pembentukan pola ataupun dasar pengambilan keputusan dalam proses prediksi penyakit diabetes.

2. METODOLOGI PENELITIAN

2.1 Kerangka Kerja Penelitian

Kerangka kerja penelitian juga biasa disebut dengan metodologi penelitian. Disini akan menggambarkan setiap tahapan proses yang dilakukan pada penelitian. Dimana proses tersebut dimulai dari identifikasi masalah sampai dengan kesimpulan. Peran dari pada kerangka kerja penelitian bertujuan untuk memudahkan bagi peneliti untuk mengetahui tahapan apa saja yang harus dilakukan pada penelitian. Adapun kerangka kerja penelitian yang terdapat pada penelitian ini dapat dilihat pada gambar 1 berikut:



Gambar 1. Kerangka Kerja Penelitian

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 5, No. 1, November 2023, Page 165-174
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v5i1.4614

2.2 Data Mining

Data mining merupakan sebuah teknik yang dipergunakan untuk mendapatkan kembali informasi yang tersimpan dari kumpulan data. Pada proses data mining melakukan pengulasan kembali atau pemrosesan kembali terhadap data hingga didapatkan sebuah informasi ataupun pola baru yang dipergunakan bagi pemilik data sebagai proses pengambilan keputusan. Dalam pelaksanaanya data mining sudah banyak diterapkan dalam berbagai macam bidang ilmu, hal tersebut dikarenakan sudah hampir seluruh bidang ilmu dimasa sekarang ini sudah melakukan proses pengolahan data[20], [21], [34].

2.3 Algoritma Naïve Bayes

Algoritma Naïve Bayes merupakan algoritma klasifikasi pada data mining. Algoritma Naïve Bayes melakukan proses klasifikasi berdasarkan dengan konsep statistika, hal tersebut didasari dengan alogirtma Naïve Bayes mengadopsi dari Teorema Bayes. Setiap kelas pada algoritma Naïve Bayes memiliki nilai probablitias tersendiri yang nantinya dapat digunakan untuk proses pengambilan keputusan. Adapun persamaan umum yang digunakan yaitu[35]–[37]:

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \quad (1)$$

2.4 Algoritma C4.5

Algoritma 4.5 merupakan bagian dari proses klasifikasi pada data mining. Pada algoritma C4.5 proses dilakukan dengan membentukan pohon keputusan dan mendapatkan rule aturan yang baru. Pembentukan akar pada pohon keputusan dari perhitungan nilai infromation dan gain yang melakukan pemecahan kelompok beradasarkan dengan kelasnya masing – masing. Adapun rumus penyelesaian dapat dilihat pada berikut[38]–[40]:

$$Entropy\ (S) = \sum_{i=0}^{n} - pi * log^2\ pi \quad (2)$$

$$Gain\ (S, A) = Entropy\ (S) - \sum_{i=1}^{n} \frac{|Si|}{|s|} * Entropy\ (Si) \quad (3)$$

2.5 Algoritma K-Nearest Neighbor (K-NN)

Algoritma K-Nearest Neighbor (K-NN) merupakan bagian dari pada teknik klasifikasi pada data mining. Dimana Algoritma K-Nearest Neighbor (K-NN) dapat melakukan pengelompokan terhadap data yang baru untuk dimasukan pada kelompok tertentu berdasarkan dengan kriteria tertentu. Proses pengelompokan yang dilakukan pada Algoritma K-Nearest Neighbor (K-NN) dengan melakukan perhitungan jarak terdekat pada data baru terhadap jarak terhadap data lama. Proses yang dilakukan pada Algoritma K-Nearest Neighbor (K-NN) berdasarkan dengan perhitungan nilai euclidean distance. Adapun rumus yang digunakan untuk proses perhitungan jarak dapat dilihat berikut[6], [18], [23]:

$$Dq = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + (a_n - b_n)^2} \quad (1)$$

2.6 Pengumpulan Data

Pengumpulan data merupakan berdasarkan dengan data historis dari penyakit diabetes sebelumnya. Pada penelitian ini proses penelitian dilakukan dengan menggunakan dataset dari penyakit diabetes https://www.kaggle.com/datasets/mathchi/diabetes-data-set merupakan link yang digunakan untuk mengakses dari dataset diabetes.

3. HASIL DAN PEMBAHASAN

3.1 Analisa Masalah

Diabetes merupakan penyakit yang sangat dekat dengan kehidupan manusia. Dimana penyakit ini dikarenakan terlalu banyaknya kadar gula didalam tubuh manusia. Penyakit diabetes ini disebabkan karena terhambatnya proses metabolisme tubuh untuk melakukan proses pencernaan terhadap gula didalam tubuh yang menyebabkan gula menumpuk pada tubuh hingga kadar yang melewati batas ketentuan. Penyakit diabetes ini bukan merupakan penyakit menular, tetapi jika terjadi pada manusia akan sangat sulit untuk dilakukan penyembuhan. Selain itu juga penyakit ini memiliki dampak yang buruk bagi tubuh serta kesehatan manusia dimana penyakit ini dapat menimbulkan komplikasi terhadap penyakit – penyakit lainnya yang terkadang menyebabkan hingga kematian. Maka dari itu perlu sejak dini untuk dilakukan prediksi terhadap seseorang untuk mengetahui apakah terdapat penyakit diabetes atau mungkin terdapat kemungkinan seseorang tersebut menderita penyakit diabetes. Prediksi dilakukan bertujuan untuk memperkirakan hasil dimasa yang akan mendatang, pada penelitian ini prediksi dilakukan dengan menggunakan pola yang terbentuk dari kriteria atribut pada penyakit. Prediksi terhadap penyakit diabetes dapat diselesaikan dengan menggunakan data mining. Data mining merupakan proses pengolahan data yang bertujuan untuk mendapatkan kembali sebuah informasi penting yang tersimpan pada

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 5, No. 1, November 2023, Page 165-174
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v5i1.4614

data. Pada data mining data yang digunakan merupakan data dimasa lampau yang tersimpan pada gudang data yang kemudian dilakukan pengolahan kembali hingga terbentuk informasi untuk pengambilan keputusan. Penyelesaian pada data mining sendiri terdapat berbagai macam cara didalamnya, salah satu adalah klasifikasi. Klasifikasi merupakan proses pengelompokan data pada kelas – kelas tertentu. Pengelompokan data pada klasifikasi berdasarkan dengan kombinasi yang terdapat pada atribut data untuk membentuk kelas – kelas tertentu. Pada penelitian ini untuk penyelesaian klasifikasi menggunakan algoritma Naïve Bayes, K-Nearest Neighbor dan C4.5. Penggunaan algoritma – algoritma tersebut bertujuan untuk melakukan perbandingan, proses perbandingan untuk mementunkan pola mana yang paling baik digunakan dalam proses prediksi penyakit diabetes. Pola yang paling baik didapatkan berdasarkan dengan hasil kinerja yang paling tinggi dari algoritma – algoritma klasifikasi tersebut.

3.2 Hasil Pengumpulan Data

Sebelum dilakukan proses penyelesaian pada penelitian diharuskan terlebih dahulu untuk mengetahui data yang akan digunakan. Dalam hal ini penelitian ini menggunakan dataset penyakit diabetes yang sudah tersedia pada link diatas. Pada dataset tersebut terdapat 8 (delapan) atribut yaitu Pregnancies, Glucose, BloodPressure, SkinThickness, Insulin, BMI, DiabetesPedigreeFunction dan Age. Selain atribut pada dataset juga terdapat kelas yang menjadi target yaitu Outcome. Pada dataset penyakit diabtes tersebut terdapat 768 record data. Adapun sampel dataset penyakit diabetes dapat dilihat pada tabel 1 berikut:

Tabel 1. Dataset Penyakit Diabetes

Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
6	148	72	35	0	33.6	0.627	50	1
1	85	66	29	0	26.6	0.351	31	0
8	183	64	0	0	23.3	0.672	32	1
1	89	66	23	94	28.1	0.167	21	0
0	137	40	35	168	43.1	2.288	33	1
5	116	74	0	0	25.6	0.201	30	0
3	78	50	32	88	31	0.248	26	1
10	115	0	0	0	35.3	0.134	29	0
2	197	70	45	543	30.5	0.158	53	1
8	125	96	0	0	0	0.232	54	1
…	…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…	…
1	93	70	31	0	30.4	0.315	23	0

Sebelum dilakukan proses penyelesaian dengan menggunakan algoritma tersebut. Perlu kiranya dilakukan tahapan preprocesing data. Tahapan preprocesing data bertujuan untuk menyesuaikan data dengan algoritma yang akan digunakan. Dalam hal ini nilai yang terdapat pada dataset harus dilakukan pengelompokan terhadap kategorikal tertentu. Adapun hasil preprocesing data dapat dilihat pada tabel 2 berikut:

Tabel 2. Hasil Preprocesing Data

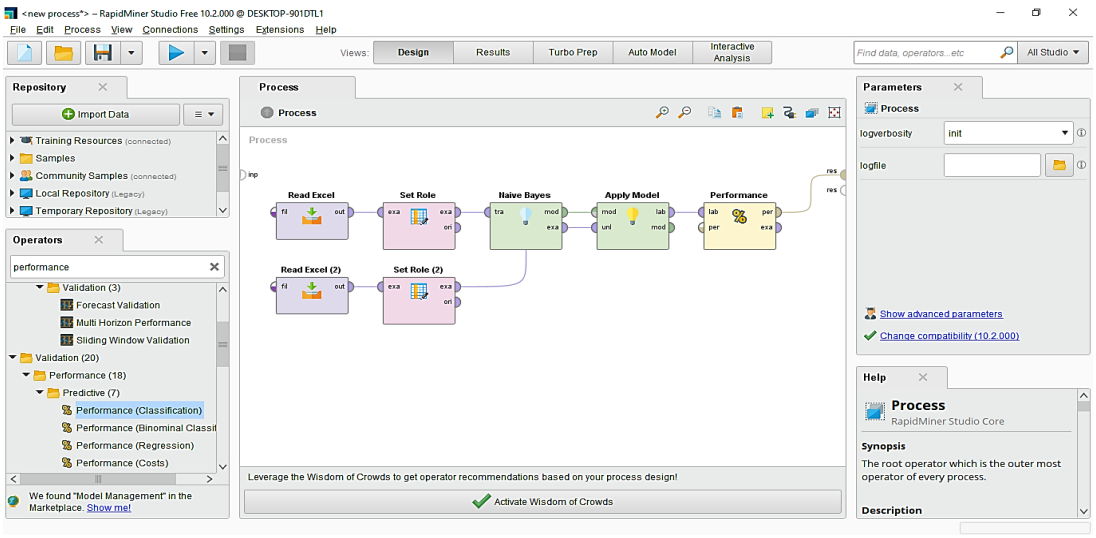
Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
Banyak	Sedang	Normal	Sedang	Sedikit	Obesitas	Tinggi	Orang Tua	Ya
Sedikit	Sedikit	Normal	Sedang	Sedikit	Overweight	Sedang	Dewasa	Tidak
Banyak	Tinggi	Normal	Tipis	Sedikit	Normal	Tinggi	Dewasa	Ya
Sedikit	Sedikit	Normal	Tipis	Sedikit	Overweight	Rendah	Remaja	Tidak
Tidak	Sedan	Normal	Sedang	Seda	Obesitas	Tinggi	Dewasa	Ya

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 5, No. 1, November 2023, Page 165-174
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v5i1.4614

Pernah	g			ng				
Sedang	Sedang	Normal	Tipis	Sedikit	Overweight	Rendah	Dewasa	Tidak
Sedikit	Sedikit	Normal	Sedang	Sedikit	Obesitas	Rendah	Dewasa	Ya
Banyak	Sedang	Normal	Tipis	Sedikit	Obesitas	Rendah	Dewasa	Tidak
Sedikit	Tinggi	Normal	Tinggi	Tinggi	Obesitas	Rendah	Orang Tua	Ya
Banyak	Sedang	Hipertensi	Tipis	Sedikit	Underweight	Rendah	Orang Tua	Ya
…	…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…	…
Sedikit	Sedikit	Normal	Sedang	Sedikit	Obesitas	Sedang	Remaja	Tidak

3.3 Hasil Pengujian Algoritma Naïve Bayes

Pengujian pertama sekali dilakukan dengan menggunakan algoritma Naïve Bayes. Pengujian dilakukan dengan menggunakan tools rapid miner studi. Dengan menggunakan tools tersebut pengujian tersebut berdasarkan operator yang telah tersebut. Adapun gambaran pengujian dapat dilihat pada gambar 2 berikut.



Gambar 2. Proses Pengujian Algoritma Naïve Bayes

Pada Gambar 2 diatas merupakan operator yang digunakan untuk melakukan pengujian algoritma Naïve Bayes pada rapidminer studio. Dari operator tersebut nantinya akan didapatkan hasil pengujian. Adapun hasil pengujian dapat dilihat berikut:

accuracy: 75.00%

	true Ya	true Tidak	class precision
pred. Ya	170	94	64.39%
pred. Tidak	98	406	80.56%
class recall	63.43%	81.20%	

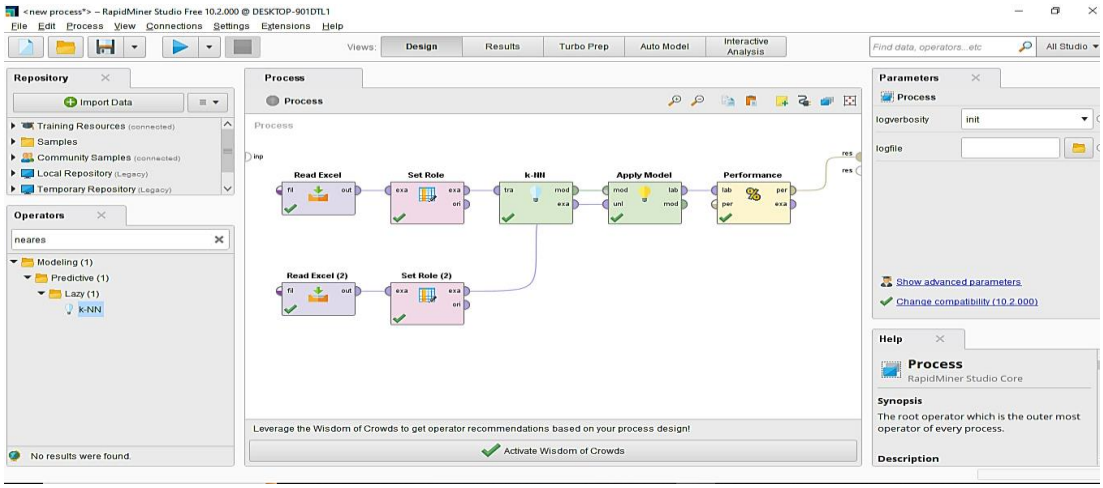
Gambar 3. Hasil Pengujian Algoritma Naïve Bayes

Dapat dilihat pada gambar 3 tersebut bahwasannya dari hasil pengujian yang dilakukan kinerja dari pada algoritma Naïve Bayes sebesar 75%.

3.4 Hasil Pengujian Algoritma K-Nearest Neighbor

Pengujian selanjutnya dilakukan dengan menggunakan algoritma K-Nearest Neighbor. Pengujian dilakukan dengan menggunakan tools rapid miner studi. Dengan menggunakan tools tersebut pengujian tersebut berdasarkan operator yang telah tersebut. Adapun gambaran pengujian dapat dilihat pada gambar 4 berikut:

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 5, No. 1, November 2023, Page 165-174
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v5i1.4614



Gambar 4. Proses Pengujian Algoritma K-Nearest Neighbor

Pada Gambar 4 merupakan operator yang digunakan untuk melakukan pengujian algoritma K-Nearest Neighbor pada rapidminer studio. Dari operator tersebut nantinya akan didapatkan hasil pengujian. Adapun hasil pengujian dapat dilihap pada gambar 5 berikut:

accuracy: 80.60%

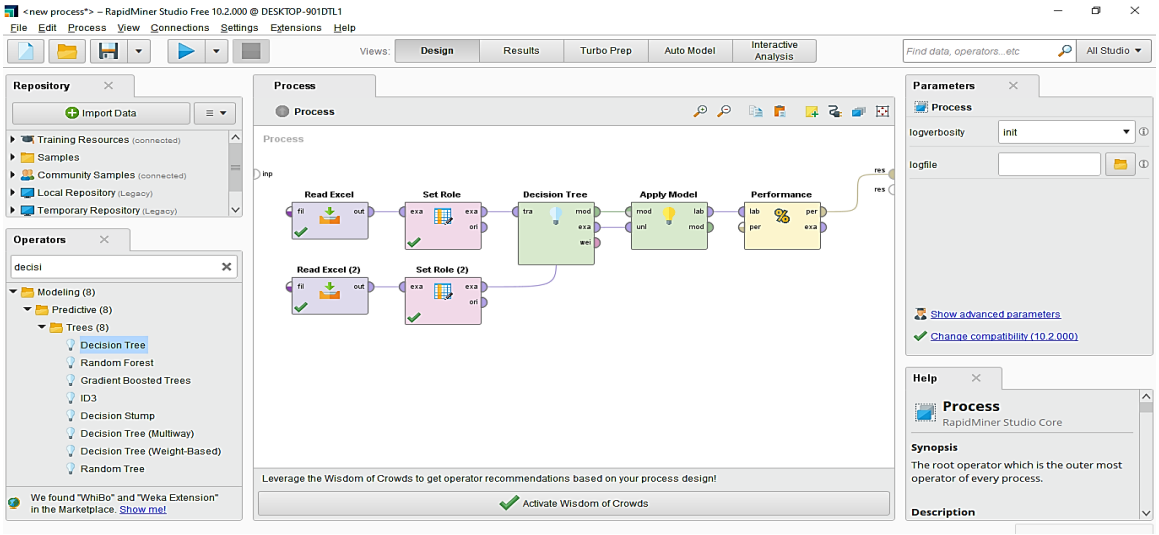
	true Ya	true Tidak	class precision
pred. Ya	178	59	75.11%
pred. Tidak	90	441	83.05%
class recall	66.42%	88.20%	

Gambar 5. Hasil Pengujian Algoritma K-Nearest Neighbor

Dapat dilihat pada gambar 5 tersebut bahwasannya dari hasil pengujian yang dilakukan kinerja dari pada algoritma K-Nearest Neighbor sebesar 80,60%.

3.5 Hasil Pengujian Algoritma C4.5

Pengujian terakhir dilakukan dengan menggunakan algoritma C4.5. Pengujian dilakukan dengan menggunakan tools rapid miner studi. Dengan menggunakan tools tersebut pengujian tersebut berdasarkan operator yang telah tersebut. Adapun gambaran pengujian dapat dilihat pada gambar 6 berikut:



Gambar 6. Proses Pengujian Algoritma C4.5

Pada Gambar 6 diatas merupakan operator yang digunakan untuk melakukan pengujian algoritma C4.5 pada rapidminer studio. Dari operator tersebut nantinya akan didapatkan hasil pengujian. Adapun hasil pengujian dapat dilihat berikut:

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 5, No. 1, November 2023, Page 165-174
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v5i1.4614

accuracy: 91.80%

	true Ya	true Tidak	class precision
pred. Ya	246	41	85.71%
pred. Tidak	22	459	95.43%
class recall	91.79%	91.80%	

Gambar 7. Hasil Pengujian Algoritma C4.5

Dapat dilihat pada gambar 7 tersebut bahwasannya dari hasil pengujian yang dilakukan kinerja dari pada algoritma C4.5 sebesar 91,80%.

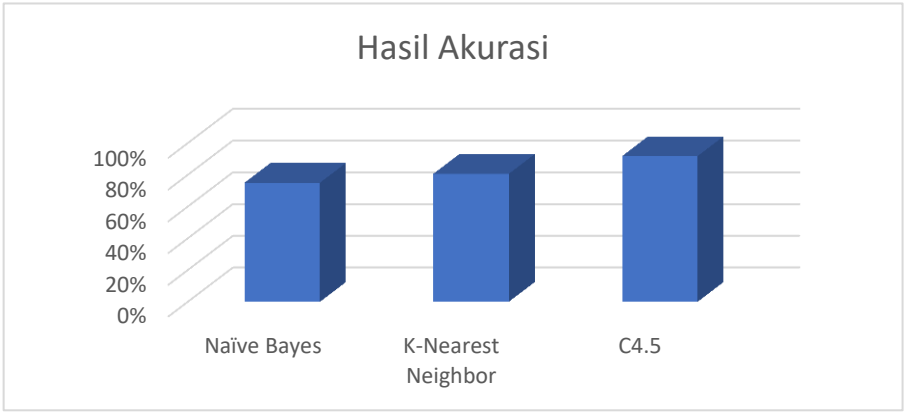
3.6 Pembahasan

Setelah dilakukan pengujian terhadap algoritma – algoritma klasifikasi maka selanjutnya dilakukan proses pembahasan untuk didapatkan hasil penelitian. Untuk hasil pengujian dari ketiga algoritma dapat dilihat pada tabel 3 berikut

Tabel 3. Hasil Pengujian Algoritma Klasifikasi

No	Alogritma	Hasil Akurasi
1	Naïve Bayes	75%
2	K-Nearest Neighbor	80,60%
3	C4.5	91,80%

Dari tabel diatas merupakan hasil proses akurasi yang didapatkan pada tahapan pengujian algoritma ataupun dari tabel tersebut dapat dibuatkan dalam bentuk gambar 8 grafik seperti berikut:



Gambar 8. Hasil Pengujian Algoritma Klasifikasi

Dari hasil proses pengujian algoirtma klasifikasi terhadap penyakit diabetes tersebut, dapat dilihat bahwasannya algoritma C4.5 mendapatkan kinerja lebih baik dibandingan dengan algoritma lainnya. Dimana algoritma C4.5 mendapatkan akurasi sebesar 91,80%. Dalam hal ini maka hasil yang didapatkan pada algoritma C4.5 dipergunakan sebagai pola model untuk melakukan prediksi terhadap penyakit diabtes.

4. KESIMPULAN

Setelah dilakukan serangkaian proses penelitian dimulai dari awal sampai dengan akhir penelitian maka selanjutnya dapat diambil kesimpulan dalam penelitian. Adapun yang menjadi kesimpulan dari penelitian yaitu proses prediksi terhadap penyakit diabetes dapat diselesaikan dengan menggunakan data mining. Pada data mining proses prediksi dilakukan dengan memanfaatkan dari pada teknik klasifikasi. Pada proses klasifikasi data mining terdapat 3 (tiga) algoritma yang digunakan yaitu Naïve Bayes, K-Nearest Neighbor dan C4.5 Dari hasil pengujian yang telah dilakukan didapatkan hasil kinerja akurasi untuk algoritma Naïve Bayes sebesar 75%, akurasi untuk algoirtma K-Nearest Neighbor sebsar 80,60% dan algoritma C4.5 sebsar 91,80%. Dalam hal ini menandakan bahwasannya algoritma C4.5 memiliki kinerja lebih baik dibandingan dengan algroritma lainnya.
Maka dari itu hasil pola yang dihasil dari algoritma C4.5 dipergunakan untuk melakukan prediksi terhadap penyakit diabtes tersebut.

REFERENCES

[1] A. Veronica Agustin, A. Voutama, U. J. Singaperbangsa Karawang HS Ronggo Waluyo, and J. Barat, “Implementasi Data Mining Klasifikasi Penyakit Diabetes Pada Perempuan Menggunakan Naïve Bayes,” *J. Mhs. Tek. Inform.*, vol. 7, no. 2, pp. 1002–1007, 2023.

ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 5, No. 1, November 2023, Page 165-174
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v5i1.4614

[2] D. U. Iswavigra, S. Defit, and G. W. Nurcahyo, "Data Mining dalam Pengelompokan Penyakit Pasien dengan Metode K-Medoids," *J. Inf. dan Teknol.*, vol. 3, pp. 181–189, 2021, doi: 10.37034/jidt.v3i4.150.
[3] F. M. Hana, "Klasifikasi Penderita Penyakit Diabetes Menggunakan Algoritma Decision Tree C4.5," *J. SISKOM-KB (Sistem Komput. dan Kecerdasan Buatan)*, vol. 4, no. 1, pp. 32–39, 2020, doi: 10.47970/siskom-kb.v4i1.173.
[4] N. Nurdiana and A. Algifari, "Studi Komparasi Algoritma ID3 Dan Algoritma Naive Bayes Untuk Klasifikasi Penyakit Diabetes Mellitus," *INFOTECH J.*, vol. 6, no. 2, pp. 18–23, 2020.
[5] A. Ridwan, "Penerapan Algoritma Naïve Bayes Untuk Klasifikasi Penyakit Diabetes Mellitus," *J. SISKOM-KB (Sistem Komput. dan Kecerdasan Buatan)*, vol. 4, no. 1, pp. 15–21, 2020, doi: 10.47970/siskom-kb.v4i1.169.
[6] M. S. Mustafa and I. W. Simpen, "Implementasi Algoritma K-Nearest Neighbor (KNN) Untuk Memprediksi Pasien Terkena Penyakit Diabetes Pada Puskesmas Manyampa Kabupaten Bulukumba," *Semin. Ilm. Sist. Inf. Dan Teknol. Inf.*, vol. VIII, no. 1, pp. 1–10, 2019, [Online]. Available: https://ejurnal.dipanegara.ac.id/index.php/sisiti/article/view/1 -10/68.
[7] N. Novianti, M. Zarlis, and P. Sihombing, "Penerapan Algoritma Adaboost Untuk Peningkatan Kinerja Klasifikasi Data Mining Pada Imbalance Dataset Diabetes," *J. Media Inform. Budidarma*, vol. 6, no. 2, p. 1200, 2022, doi: 10.30865/mib.v6i2.4017.
[8] M. A. Wiratama and W. M. Pradnya, "Optimasi Algoritma Data Mining Menggunakan Backward Elimination untuk Klasifikasi Penyakit Diabetes," *J. Nas. Pendidik. Tek. Inform.*, vol. 11, no. 1, p. 1, 2022, doi: 10.23887/janapati.v11i1.45282.
[9] Normah, B. Rifai, S. Vambudi, and R. Maulana, "Analisa Sentimen Perkembangan Vtuber Dengan Metode Support Vector Machine Berbasis SMOTE," *J. Tek. Komput. AMIK BSI*, vol. 8, no. 2, pp. 174–180, 2022, doi: 10.31294/jtk.v4i2.
[10] B. A. Candra Permana and I. K. Dewi Patwari, "Komparasi Metode Klasifikasi Data Mining Decision Tree dan Naïve Bayes Untuk Prediksi Penyakit Diabetes," *Infotek J. Inform. dan Teknol.*, vol. 4, no. 1, pp. 63–69, 2021, doi: 10.29408/jit.v4i1.2994.
[11] S. Ucha Putri, E. Irawan, and F. Rizky, "Implementasi Data Mining Untuk Prediksi Penyakit Diabetes Dengan Algoritma C4.5," *KESATRIA J. Penerapan Sist. Inf. (Komputer Manajemen)*, vol. 2, no. 1, pp. 39–46, 2021.
[12] F. Aris and Benyamin, "Penerapan Data Mining untuk Identifikasi Penyakit Diabetes Melitus dengan Menggunakan Metode Klasifikasi," *Router Res.*, vol. 1, no. 1, pp. 1–6, 2019, [Online]. Available: https://www.ejournal.stipwunaraha.ac.id/index.php/router/article/view/313.
[13] D. Marlina and M. Bakri, "Penerapan Data Mining Untuk Memprediksi Transaksi Nasabah Dengan Algoritma C4.5," *J. Teknol. dan Sist. Inf.*, vol. 2, no. 1, pp. 23–28, 2021.
[14] D. P. Utomo and B. Purba, "Penerapan Datamining pada Data Gempa Bumi Terhadap Potensi Tsunami di Indonesia," *Pros. Semin. Nas. Ris. Inf. Sci.*, vol. 1, no. 1, pp. 846–853, 2019.
[15] R. Takdirillah, "Penerapan Data Mining Menggunakan Algoritma Apriori Terhadap Data Transaksi Sebagai Pendukung Informasi Strategi Penjualan," *Edumatic J. Pendidik. Inform.*, vol. 4, no. 1, pp. 37–46, 2020, doi: 10.29408/edumatic.v4i1.2081.
[16] M. Nizam Fadli, I. Sudahri Damanik, E. Irawan, S. Tunas Bangsa, and S. Utara, "Penerapan Metode Naive Bayes Dalam Menentukan Tingkat Kenyamanan Pada Rumah Sakit Terhadap Pasien," *KLIK Kaji. Ilm. Inform. dan Komput. ,* vol. 2, no. 3, pp. 117–122, 2021, [Online]. Available: https://djournals.com/klik.
[17] Tugiman, Lily Damayanti, Alexius Hendra Gunawan, and Samuel Ryon Elkana, "Prediksi Penggunaan Obat Peserta Jaminan Kesehatan Nasional Menggunakan Algoritma Naïve Bayes Classifier," *J. Appl. Comput. Sci. Technol.*, vol. 3, no. 1, pp. 144–150, 2022, doi: 10.52158/jacost.v3i1.295.
[18] S. Cumel, David Zamri, Rahmaddeni, "Perbandingan Metode Data Mining untuk Prediksi Banjir Dengan Algoritma Naïve Bayes dan KNN," *SENTIMAS Semin. Nas. Penelit. dan …*, pp. 40–48, 2022, [Online]. Available: https://journal.irpi.or.id/index.php/sentimas/article/view/353%0Ahttps://journal.irpi.or.id/index.php/sentimas/article/download/353/132.
[19] N. S. Fauziah and R. D. Dana, "Implementasi Algoritma Naive bayes dalam Klasifikasi Status Kesejahteraan Masyarakat Desa Gunungsari," *Blend Sains J. Tek.*, vol. 1, no. 4, pp. 295–305, 2023, doi: 10.56211/blendsains.v1i4.234.
[20] I. Nurjanah, J. Karaman, I. Widaningrum, D. Mustikasari, and S. Sucipto, "Penggunaan Algoritma Naïve Bayes Untuk Menentukan Pemberian Kredit Pada Koperasi Desa," *Explorer (Hayward).*, vol. 3, no. 2, pp. 77–87, 2023.
[21] Q. A'yuniyah, W. Elvira, N. Nazira, I. Ambarani, S. F. Intan, and D. Ramadhani, "Analisa Algoritma Naïve Bayes Classifier (NBC) Untuk Prediksi Penjualan Alat Kesehatan," *IJIRSE Indones. J. Inform. Res. Softw. Eng. J.*, vol. 3, no. 2, pp. 119–126, 2023.
[22] Y. Ambar, Kusrini, and Henderi, "Penerapan Algoritma K-Nearest Neighbour Dalam Menentukan Pembinaan Koperasi Kabupaten Kotawaringin Timur," *Citec J.*, vol. 5, no. 3, pp. 232–241, 2019.
[23] A. Y. Muniar, P. Pasnur, and K. R. Lestari, "Penerapan Algoritma K-Nearest Neighbor pada

ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 5, No. 1, November 2023, Page 165-174
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v5i1.4614

Pengklasifikasian Dokumen Berita Online," *Inspir. J. Teknol. Inf. dan Komun.*, vol. 10, no. 2, p. 137, 2020, doi: 10.35585/inspir.v10i2.2570.

[24] M. F. A. Sularno, W. Wiyanto, D. Ardiatma, and A. T. Zy, "Penerapan Algoritma K-Nearest Neighbor pada Klasifikasi Penyakit Jantung," *J. Comput. Syst. Informatics*, vol. 4, no. 4, pp. 850–860, 2023, doi: 10.47065/josyc.v4i4.4071.

[25] I. S. Muallif, H. Budiman, and N. Ransi, "Penerapan Data Mining untuk Prediksi Pergerakan Harga Saham Menggunakan Algoritma K-Nearest Neighbor," in *PROSIDING SEMINAR NASIONAL PEMANFAATAN SAINS DAN TEKNOLOGI INFORMASI*, 2023, vol. 1, no. 1, pp. 297–306.

[26] R. Rahmadini, Enjel Erika LorencisLubis, Aji Priansyah, Yolanda R.W.N, and Tuti Meutia, "Penerapan Data Mining Untuk Memprediksi Harga Bahan Pangan Di Indonesia Menggunakan Algoritma K-Nearest Neighbor," *J. Mhs. Akunt. Samudra*, vol. 4, no. 4, pp. 223–235, 2023, doi: 10.33059/jmas.v4i4.7074.

[27] R. Bahtiar, "Implementasi Data Mining Untuk Prediksi Penjualan Kusen Terlaris Menggunakan Metode K-Nearest Neighbor," *J. Inform. MULTI*, vol. 1, no. 3, pp. 200–214, 2023, [Online]. Available: https://jurnal.publikasitecno.id/index.php/jim203.

[28] Asmira, "Penerapan Data Mining untuk Mengklasifikasi Pola Nasabah Menggunakan Algoritma C4,5 pada Bank BRI Unit Andounohu Kendari," *J. Sist. Komput. dan Sist. Inf.*, vol. 1, no. 1, pp. 22–28, 2019, [Online]. Available: http://ejournal.stipwunaraha.ac.id/index.php/router.

[29] K. Aidi Saputra, J. Tata Hardinata, M. Ridwan Lubis, S. Retno Andani, and I. Syahputra Saragih, "Klasifikasi Algoritma C4.5 Dalam Penerapan Tingkat Kepuasan Siswa Terhadap Media Pembelajaran Online," *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 1, no. 3, pp. 113–118, 2020, [Online]. Available: https://djournals.com/klik.

[30] T. Novika, P. Poningsih, H. Okprana, A. P. Windarto, and H. Siahaan, "Penerapan Data Mining Klasifikasi Tingkat Pemahaman Siswa Pada Pelajaran Matematika," *J. Media Inform. Budidarma*, vol. 5, no. 1, p. 9, 2021, doi: 10.30865/mib.v5i1.2498.

[31] F. Pirmansyah and T. Wahyudi, "IMPLEMENTASI DATA MINING MENGGUNAKAN ALGORITMA C4.5 UNTUK PREDIKSI EVALUASI ANGGOTA SATUAN PENGAMANAN STUDI KASUS PT. YIMM PULOGADUNG," *J. Indones. Manaj. Inform. dan Komun.*, vol. 4, no. 2, pp. 540–551, 2023.

[32] S. Mulyanda, S. Defit, and Sumijan, "Analisis Data Mining Menggunakan Algoritma C4.5 Untuk Prediksi Harga Pasar Mobil Bekas," *J. KomtekInfo*, vol. 10, no. 1, pp. 116–121, 2023, doi: 10.35134/komtekinfo.v10i3.427.

[33] T. Widiastuti, K. Karsa, and C. Juliane, "Evaluasi Tingkat Kepuasan Mahasiswa Terhadap Pelayanan Akademik Menggunakan Metode Klasifikasi Algoritma C4.5," *Technomedia J.*, vol. 7, no. 3, pp. 364–380, 2022, doi: 10.33050/tmj.v7i3.1932.

[34] E. Budiarto, R. Rino, S. Hariyanto, and D. Susilawati, "Penerapan Data Mining Untuk Rekomendasi Beasiswa Pada SD Maria Mediatrix Menggunakan Algoritma C4.5," *J. ALGOR*, vol. 3, no. 2, pp. 23–34, 2022, doi: 10.31253/algor.v3i2.1019.

[35] E. Febriyani and H. Februariyanti, "Analisis Sentimen Terhadap Program Kampus Merdeka Menggunakan Naive Bayes Di Twitter," *J. TEKNO KOMPAK*, vol. 17, no. 2, pp. 25–38, 2022.

[36] N. Agustina and M. Hermawati, "Implementasi Algoritma Naïve Bayes Classifier untuk Mendeteksi Berita Palsu pada Sosial Media," *Fakt. Exacta*, vol. 14, no. 4, pp. 1979–276, 2021, doi: 10.30998/faktorexacta.v14i4.11259.

[37] A. I. Tanggraeni and M. N. N. Sitokdana, "Analisis Sentimen Aplikasi E-Government pada Google Play Menggunakan Algoritma Naïve Bayes," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 9, no. 2, pp. 785–795, 2022, doi: 10.35957/jatisi.v9i2.1835.

[38] R. R. Andarista and A. Jananto, "Penerapan Data Mining Algoritma C4 . 5 Untuk Klasifikasi Hasil Pengujian Kendaraan Bermotor," vol. 16, no. 2, pp. 29–43.

[39] I. Romli and A. T. Zy, "Penentuan Jadwal Overtime Dengan Klasifikasi Data Karyawan Menggunakan Algoritma C4.5," *J. Sains Komput. Inform. (J-SAKTI*, vol. 4, no. 2, pp. 694–702, 2020.

[40] N. Azwanti and E. Elisa, "Analisa Kepuasan Konsumen Menggunakan Algoritma C4.5," *Pros. Semin. Nas. Ilmu Sos. dan Teknol.*, no. 3, pp. 126–131, 2020.