cek klasifikasi adidas

by UNIVERSITAS AHMAD DAHLAN 12

Submission date: 15-Mar-2025 04:10PM (UTC+0700)
Submission ID: 2614498662
File name: adidas.pdf (1.28M)
Word count: 6305
Character count: 38768

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 5, Nomor 2, Desember 2023
e-ISSN 2685-998X
DOI 10.30865/json.v5i2.6124

Perbandingan Algoritma Naive Bayes, Decision Tree, dan KNN untuk Klasifikasi Produk Populer Adidas US dengan Confusion Matrix

Lazuardi Firdaus, Tedy Setiadi*
Fakultas Teknologi Industri, Informatika, Universitas Ahmad Dahlan, Yogyakarta, Indonesia
Email: [1]lazuardi1800018319@webmail.uad.ac.id, [2,*]tedy.setiadi@tif.uad.ac.id
Email Penulis Korespondensi: tedy.setiadi@tif.uad.ac.id
Submitted: **04/05/2023**; Accepted: **16/10/2023**; Published: **22/12/2023**

Abstrak–Adidas America, Inc. (nama alternatif: Adidas US) merupakan perusahaan yang memproduksi sepatu, pakaian, dan aksesoris sebagai anak perusahaan dari Adidas AG yang dikenal di seluruh dunia dengan tanda identitas berupa tiga garis pada produknya. Kepopuleran produk sangat penting untuk peningkatan penjualan terutama pada produk yang sering dibeli, mendapat penilaian positif dari pelanggan, dan banyak diulas oleh pelanggan. Pada kasus ini, terdapat produk-produk Adidas US yang kepopulerannya masih belum diketahui. Karena alasan tersebut, produk dari Adidas US yang populer akan diklasifikasikan sebagai kebutuhan bisnis dari Adidas US. Algoritma klasifikasi yang akan digunakan untuk klasifikasi produk populer adalah Naive Bayes, Decision Tree, dan KNN untuk klasifikasi popularitas produk Adidas di Amerika Serikat dengan menggunakan metode CRISP-DM pada dataset. Proses data mining dilakukan untuk menemukan pola dalam kumpulan data dengan tahapan pemahaman bisnis, pra-pemrosesan data, dan pemodelan klasifikasi dengan tiga algoritma yang berbeda. Hasil dari ketiga algoritma diuji dengan confusion matrix dan nilai akurasi tertinggi dicapai oleh Decision Tree dengan 92,42%, sehingga algoritma tersebut dianggap sebagai yang terbaik untuk klasifikasi produk populer Adidas US.

Kata Kunci: Perbandingan; Klasifikasi; Naive Bayes; Decision Tree; KNN

Abstract– Adidas America, Inc. (also known as Adidas US) is a company that produces shoes, clothing, and accessories as a subsidiary of Adidas AG, which is known worldwide for its trademark three stripes on its products. Product popularity is very important in increasing sales, especially for products that are frequently purchased, positively reviewed by customers, and reviewed by customers. In this case, there are Adidas US products whose popularity is still unknown. Therefore, in this case, popular Adidas US products will be classified as business needs of Adidas US. The classification algorithm used to classify popular products is Naive Bayes, Decision Tree, and KNN for classifying the popularity of Adidas products in the United States using the CRISP-DM method on the dataset. The data mining process is performed to discover patterns in the data set with the stages of business understanding, data pre-processing, and classification modeling with three different algorithms. The results of the three algorithms are tested with a confusion matrix, and the highest accuracy value is achieved by Decision Tree with 92.42%, making it the best algorithm for classifying popular Adidas US products.

Keywords: Comparison; Classification; Naive Bayes; Decision Tree; KNN

1. PENDAHULUAN

Penambangan data atau data mining adalah suatu proses untuk mengambil data dan menemukan pola pada kumpulan data dengan metode ilmu data. Dalam melakukan penambangan data, studi kasus atau permasalahannya harus dibuat terlebih dahulu untuk mengetahui data apa saja yang dibutuhkan. Kepopuleran produk sangat penting dalam meningkatkan penjualan terutama pada produk yang sering dibeli, mendapat penilaian positif dari pelanggan, dan banyak diulas oleh pelanggan. Oleh karena itu, tujuan utama dari penelitian ini adalah membandingkan tiga algoritma klasifikasi: Naïve Bayes, Decision Tree, dan KNN, guna mendapatkan hasil klasifikasi data produk populer dari Adidas US (Adidas America, Inc.). Dalam perbandingan algoritma ini, akan ditentukan algoritma terbaik untuk melakukan klasifikasi pada data produk populer Adidas US. Hasil klasifikasi data yang diharapkan adalah klasifikasi data dari kategori produk populer dengan nilai data yaitu “No" (tidak populer) dan “Yes” (populer).

Menurut Setio dkk. [1], klasifikasi merupakan suatu teknik dalam data mining yang mengalokasikan data ke dalam kelompok atau kelas yang telah ditentukan. Dalam pandangan Dunham yang dijelaskan pada Setio dkk. [1], klasifikasi juga dikenal sebagai metode supervised learning yang mengharuskan adanya data latih yang telah diberi label sehingga dapat menghasilkan aturan untuk mengklasifikasikan data uji ke dalam kelompok atau kelas yang telah ditentukan.

Dengan kata lain, klasifikasi data digunakan dalam membuat klasifikasi pada kelompok data berdasarkan data yang sudah ada pada data latih. Dalam penelitian ini juga memberikan penjelasan tentang penerapan CRISP-DM (Cross-Industry Standard Process of Data Mining) yang memberikan uraian tentang bagaimana cara mengaplikasikan data mining dengan standar industri untuk membuat data dapat diproses dengan kualitas yang baik. Berdasarkan Schröer dkk. [2], CRISP-DM adalah model proses standar yang umum digunakan dalam proyek data mining. Peréz dan Calisto [3] juga menyatakan bahwa metodologi CRISP-DM dipilih untuk menstrukturkan proyek, yang mana dirancang khusus untuk proyek penambangan data lintas industri.

Berikut ini beberapa penelitian mengenai perbandingan algoritma klasifikasi. Penelitian oleh Iskandar dan Nataliani [4] membahas mengenai klasifikasi analisis sentimen pada gadget menggunakan data media sosial dengan memfokuskan pada empat aspek: harga, desain, spesifikasi, dan citra merek. Metode CRISP-DM digunakan untuk mengolah data, dan dalam tahap modelling, klasifikasi dilakukan menggunakan algoritma Naive

Copyright © 2023 **Lazuardi Firdaus**, Page 185
This Journal is licensed under a Creative Commons Attribution 4.0 International License

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 5, Nomor 2, Desember 2023
e-ISSN 2685-998X
DOI 10.30865/json.v5i2.6124

Bayes, SVM, dan k-NN untuk mendapatkan wawasan dan nilai penting dari data. Hasil penelitian menunjukkan bahwa model klasifikasi SVM memiliki kinerja terbaik. Akurasi rata-rata SVM mencapai 96.43% untuk empat aspek yang diuji: desain sebesar 94.40%, harga sebesar 97.44%, spesifikasi sebesar 96.22%, dan citra merek sebesar 97.63%.

Penelitian oleh Prayoga dkk. [5] mengulas perbandingan algoritma dalam memprediksi kesuksesan start-up dengan menggunakan dataset sekunder. Penelitian tersebut menyoroti pentingnya peran start-up dalam mendorong pertumbuhan ekonomi melalui penciptaan peluang kerja. Dalam penelitian ini, proses klasifikasi dilakukan pada 923 data start-up menggunakan ketiga algoritma, yaitu Decision Tree, kNN, dan Naive Bayes. Berdasarkan hasil pengujian menggunakan metode cross-validation dan T-test, algoritma Decision Tree terbukti menjadi yang paling akurat dengan tingkat akurasi sebesar 79,29%, melebihi kNN dengan akurasi sebesar 66,69% dan Naïve Bayes dengan akurasi sebesar 64,21%.

Penelitian dari Tri Wiyanti dan Ainurrohmah [6] menganalisis kinerja berbagai algoritma klasifikasi dalam mengklasifikasikan zona daerah risiko COVID-19 di Indonesia. Penelitian ini mengimplementasikan klasifikasi data zona risiko COVID-19 dengan kelas "rendah", "sedang", dan "tinggi". Hasil penelitian menunjukkan bahwa algoritma Decision Tree memiliki kinerja terbaik dengan tingkat akurasi 87% dalam waktu eksekusi sebesar 2 detik. Walaupun algoritma Naïve Bayes memiliki waktu eksekusi tercepat, namun performanya tidak begitu tinggi. Dalam pola klasifikasi zona daerah risiko COVID-19, ditemukan 20 aturan, dan atribut seperti "aktif", "CR", "CFR", "laju insidensi", dan "positif" terbukti memiliki pengaruh dalam mengklasifikasikan zona risiko COVID-19, sementara atribut "sembuh" tidak memiliki pengaruh yang signifikan.

Penelitian oleh Anam dan Santoso [7] menjelaskan bahwa fokus penelitian ini adalah untuk melakukan identifikasi algoritma dengan kinerja terbaik dalam melakukan klasifikasi penerima beasiswa. Berdasarkan hasil penelitian, algoritma C4.5 menunjukkan akurasi yang lebih tinggi dibandingkan dengan Naive Bayes. Evaluasi model klasifikasi menunjukkan tingkat akurasi sebesar 96,4% untuk model C4.5 dengan waktu eksekusi sebesar 0 detik, dan 95,11% untuk model Naive Bayes dengan waktu eksekusi sebesar 0 detik. Temuan ini mengindikasikan bahwa algoritma C4.5 memiliki kinerja yang lebih baik dalam mengklasifikasikan penerima beasiswa daripada algoritma Naive Bayes.

Penelitian yang dilakukan oleh Kurniawan [8] bertujuan untuk membandingkan kinerja algoritma Naive Bayes dan C4.5 dalam klasifikasi data mining dari empat studi kasus yang berbeda, yaitu penentuan penerimaan Kartu Indonesia Sehat, penentuan pengajuan kartu kredit di sebuah bank, penentuan usia kelahiran, dan penentuan kelayakan calon anggota kredit pada koperasi. Pengujian data dilakukan menggunakan algoritma Naive Bayes dan C4.5, dan hasilnya akan diimplementasikan dalam aplikasi khusus untuk manajemen klasifikasi pada keempat studi kasus tersebut. Penelitian ini menyimpulkan bahwa untuk menentukan algoritma klasifikasi terbaik dalam sebuah studi kasus, perlu mempertimbangkan kriteria, variabel, dan jumlah data yang digunakan dalam kasus tersebut.

Beberapa penelitian sebelumnya telah mengevaluasi perbandingan algoritma klasifikasi dan penggunaan CRISP-DM dalam penambangan data. Namun, penelitian sebelumnya ini juga memiliki banyak metode dan hasil data mining yang bisa dianggap cukup kompleks dan rumit untuk dipahami. Untuk itu, penelitian ini bertujuan untuk membuat klasifikasi data yang mudah untuk dipahami, serta juga untuk membuat perbandingan algoritma klasifikasi yang populer, yaitu Naive Bayes, Decision Tree, dan KNN, dengan menggunakan metodologi CRISP-DM untuk mengetahui hasil klasifikasi dari produk populer Adidas US.

Algoritma Naive Bayes, Decision Tree, dan KNN adalah algoritma yang populer dalam klasifikasi data pada data mining dan telah terbukti memberikan hasil yang baik. Hasil akhir yang akan dicapai pada penelitian ini adalah mendapatkan algoritma terbaik untuk klasifikasi produk populer Adidas US, serta hasil lainnya adalah mengetahui tingkat akurasi dari model klasifikasi untuk mendapatkan klasifikasi produk populer tersebut.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

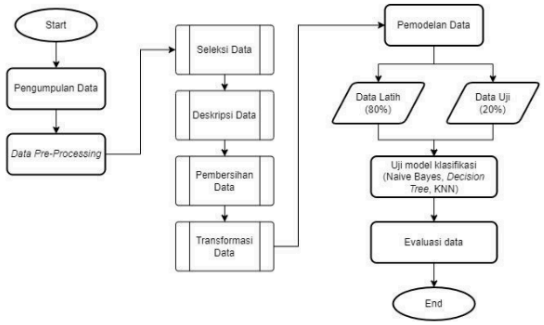
Untuk melakukan tahapan penelitian ini, model CRISP-DM (Cross-Industry Standard Process for Data Mining) akan digunakan sebagai prosedur untuk memproses data dan membuat pengujian model. Hasilnya akan menjadi klasifikasi yang akan dicapai dan dibuatkan perbandingan dari algoritma Naive Bayes, Decision Tree, dan KNN. Dalam penelitian ini, tahap Deployment yang digunakan untuk membangun sistem atau aplikasi implementasi data mining tidak digunakan. Berikut ini merupakan gambar dari model CRISP-DM pada data mining yang menjadi dasar dalam menghasilkan pola data untuk perbandingan algoritma dalam klasifikasi produk populer Adidas US.

Copyright © 2023 **Lazuardi Firdaus**, Page 186
This Journal is licensed under a Creative Commons Attribution 4.0 International License

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 5, Nomor 2, Desember 2023
e-ISSN 2685-998X
DOI 10.30865/json.v5i2.6124

Gambar 1. Model tahapan CRISP-DM

Berdasarkan gambar 1 ini, model CRISP-DM digunakan untuk mendapatkan data yang sesuai dengan kebutuhan penelitian. Tahapan penelitian juga menerapkan model CRISP-DM ini, sehingga hasil penambangan data akan menyesuaikan standar dari model ini. Berikut adalah alur diagram yang digunakan pada tahapan penelitian ini.



Gambar 2. Tahapan Penelitian

Dengan mengikuti tahapan penelitian yang dijelaskan pada gambar 2, maka setelah data sudah dikumpulkan, maka dilakukan data pre-processing (pra-pemrosesan data) untuk diolah dengan tahap seleksi data, deskripsi data, pembersihan data, dan transformasi data. Berdasarkan apa yang dijelaskan oleh Raschka [9], bahwa data yang sudah dipersiapkan akan dipisah menjadi data latih dan data uji, kemudian proporsi data dibuat berdasarkan rasio data yang umumnya adalah 60/40, 70/30, atau 80/20. Dengan penjelasan tersebut, maka data dibagi menjadi data latih dan data uji dengan rasio 80% : 20% yang dianggap rasio yang optimal untuk menghasilkan model data. Kemudian, pemodelan dilakukan dengan algoritma klasifikasi Naive Bayes, Decision Tree, dan KNN. Data yang sudah dimodelkan akan dievaluasi untuk hasil klasifikasinya.

2.2 Business Understanding

Tahap business understanding (pemahaman bisnis) bertujuan untuk memfokuskan tujuan dan kebutuhan, apakah itu untuk penelitian, bisnis, atau keperluan lain. Pada tahap ini, beberapa kegiatan yang dilakukan meliputi pemahaman terhadap kebutuhan dan tujuan dari sudut pandang bisnis, mengartikan pengetahuan menjadi definisi masalah pada data mining, dan merancang rencana dan strategi untuk mencapai tujuan data mining [10]. Langkah-langkah kunci pemahaman bisnis termasuk menentukan tujuan bisnis dan tujuan penambangan data, penilaian situasi, dan perencanaan proyek [11].

Pada tahap ini, permasalahan bisnis yang dihadapi adalah tidak adanya kategori produk populer untuk dijadikan acuan dalam penjualan produk Adidas US dan dataset yang digunakan belum memiliki atribut yang dibutuhkan dan diseleksi untuk melakukan proses klasifikasi dengan algoritma Naive Bayes, Decision Tree, dan KNN. Tujuan utama dari penelitian ini adalah untuk menguji algoritma klasifikasi pada data produk populer Adidas US dan membandingkan hasilnya untuk melihat akurasi dari klasifikasi produk populer tersebut. Tujuan analisisnya adalah menemukan hasil klasifikasi algoritma dengan akurasi yang tinggi serta membandingkan performa dari ketiga algoritma klasifikasi tersebut.

Copyright © 2023 **Lazuardi Firdaus**, Page 187
This Journal is licensed under a Creative Commons Attribution 4.0 International License

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 5, Nomor 2, Desember 2023
e-ISSN 2685-998X
DOI 10.30865/json.v5i2.6124

2.3 Data Understanding

Data Understanding (pemahaman data) merupakan tahapan penting dalam memahami kebutuhan data yang akan diproses. Setelah tahap pemahaman bisnis, fokus utama pada tahapan pemahaman data adalah untuk memperoleh data awal melalui pengumpulan data dan mempersiapkannya pada tahap persiapan data. Marzukhi, dkk. [12] memberikan penjelasan bahwa untuk fase ini, informasi data yang berkaitan akan diambil dan dikumpulkan untuk dianalisis supaya data dapat dipahami, khususnya untuk pola tersembunyi dan data yang belum terlihat.

2.3 Data Preparation

Data Preparation (persiapan data) merupakan tahapan untuk memproses data menjadi bentuk yang sesuai dengan kebutuhan yang sudah dipaparkan pada tahap business understanding dan data understanding. Tahapan untuk menghasilkan data yang dipersiapkan untuk pemodelan data umumnya menggunakan pra-pemrosesan data (data pre-processing) sebagai tahap dalam mempersiapkan data. Setelah pengumpulan data, langkah persiapan data dilakukan dengan mengidentifikasi data, memilih data yang relevan, membersihkan data, dan menyusunnya ke dalam format yang diinginkan [13].

2.4 Modelling

Modelling atau pemodelan data merupakan suatu proses atau teknik untuk membuat model matematis atau statistik yang dapat digunakan untuk memahami, menganalisis, dan memprediksi perilaku atau pola data. Pengerjaan pemodelan dilakukan dengan cara menentukan teknik-teknik data mining yang digunakan, menentukan alat-alat untuk data mining, menentukan algoritma data mining yang digunakan, dan mengatur parameter model dengan nilai-nilai optimal [14]. Proses modelling pada penelitian ini menggunakan aplikasi perangkat lunak RapidMiner versi 10.1. Pada tahap ini, model yang digunakan adalah model algoritma klasifikasi Naive Bayes, Decision Tree, dan KNN yang digunakan untuk klasifikasi produk populer Adidas US.

2.4.1 Naive Bayes

Naive Bayes merupakan suatu metode statistik yang digunakan untuk melakukan inferensi induksi dalam masalah klasifikasi. Metode ini menggunakan probabilitas yang berdasarkan pada Teorema Bayes. Pertama-tama, perlu untuk mengulas konsep dasar dan definisi pada Teorema Bayes, dan selanjutnya mengaplikasikan teorema ini untuk melakukan klasifikasi dalam data mining [15]. Algoritma Naive Bayes memiliki persamaan sebagai berikut:

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

Keterangan :

X	=	Data dengan class yang belum diketahui (target kelas/label)
H	=	Hipotesis data dari suatu class data (kelas data) yang spesifik
P(H|X)	=	Probabilitas hipotesis H berdasarkan kondisi X (posterior prob.)
P(X|H)	=	Probabilitas X berdasarkan kondisi dari H (likelihood)
P(H)	=	Probabilitas hipotesis H (prior prob.)
P(X)	=	Probabilitas dari X (evidence)

2.4.2 Decision Tree

Decision Tree adalah sebuah algoritma yang menggunakan struktur pohon keputusan untuk melakukan analisis data dan membuat keputusan berdasarkan informasi dari data. Menurut Ye di dalam Sutoyo [16], algoritma pohon keputusan ini bertujuan untuk memahami pola klasifikasi dan prediksi dari data, serta menunjukkan hubungan antara variabel atribut x dan variabel target y dalam bentuk struktur pohon. Hafizan dan Putri [17] menjelaskan bahwa pohon keputusan dapat digunakan untuk mengelompokkan data besar menjadi kelompok-kelompok data yang lebih kecil dengan mengikuti aturan-aturan keputusan tertentu. Algoritma Decision Tree menggunakan konsep entropy dan gain sebagai persamaannya. Berikut adalah penjelasan mengenai kedua konsep tersebut.

a. Entropy

Entropy merupakan ukuran yang menunjukkan tingkat ketidakpastian atau variasi keputusan terhadap nilai atribut tertentu dalam data. Semakin tinggi nilai entropy, semakin tinggi tingkat ketidakpastian atau perbedaan keputusan yang ada, dan ini akan mempengaruhi perhitungan gain. Persamaan untuk menghitung entropy adalah sebagai berikut:

$$\text{Entropy}(S) = \sum_{i=1}^{n} pi * log_2 pi \quad (2)$$

Keterangan:

S	=	Himpunan kasus
n	=	Jumlah dari atribut A
pi	=	Proporsi dari S_i terhadap S, bentuk alternatifnya adalah $\frac{\|S_i\|}{\|S\|}$
$\log_2$	=	logaritma basis 2

Copyright © 2023 **Lazuardi Firdaus**, Page 188
This Journal is licensed under a Creative Commons Attribution 4.0 International License

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 5, Nomor 2, Desember 2023
e-ISSN 2685-998X
DOI 10.30865/json.v5i2.6124

Entropy digunakan dalam pembangunan pohon keputusan untuk memahami apakah suatu atribut memiliki data yang memerlukan banyak cabang atau tidak. Ketika entropy suatu atribut adalah 0, cabang atribut tersebut berakhir pada nilai dari atribut itu sendiri. Namun, jika nilai entropy mencapai maksimum yaitu 1, cabang atribut tersebut memerlukan banyak cabang karena tingkat ketidakpastian data yang tinggi.

b. Gain

Gain merupakan salah satu metode yang digunakan untuk memilih atribut pada setiap simpul dalam pohon keputusan. Fungsi dari gain adalah mengukur sejauh mana pengaruh suatu atribut terhadap keputusan atau efektivitas variabel dalam membagi data menjadi kelas-kelas. Selain itu, gain juga sering disebut sebagai Information Gain karena menggunakan informasi yang diperlukan dalam perhitungannya. Berikut adalah persamaan untuk menghitung gain:

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

Keterangan:
S = Himpunan kasus
A = Atribut
n = Jumlah dari atribut A
$|S_i|$ = Jumlah kasus pada partisi ke-i
$|S|$ = Jumlah kasus dalam S

Setio dkk. [18] menjelaskan bahwa dalam algoritma Decision Tree, nilai gain berfungsi untuk menentukan variabel mana yang akan menjadi simpul dalam suatu pohon keputusan. Variabel yang memiliki gain tertinggi akan dijadikan simpul pada pohon keputusan tersebut.

2.4.3 K-NN

Dalam penelitian Djamsi dkk. [19], metode klasifikasi K-Nearest Neighbors (KNN) digunakan, yang didasarkan pada peninjauan objek atau tetangga terdekat. KNN mengukur jarak data dengan mengidentifikasi data terdekat dari data baru. Menurut Widua Gata dalam penelitian Nikmatul dkk. [20], algoritma KNN sederhana dalam pengoperasiannya, menghitung jarak terpendek dari data uji ke data latih untuk menentukan kelas dari data tersebut. Setelah mengumpulkan data dalam kelompok k tertentu, kelas mayoritas dari data tersebut diambil untuk dijadikan sebagai kelas prediksi dari data uji. Data yang digunakan dalam algoritma KNN harus bersifat numerik, karena perhitungannya memerlukan normalisasi untuk mempertahankan keseimbangan jarak data. Karena algoritma KNN membutuhkan konsep normalisasi data dan perhitungan euclidean distance, berikut adalah penjelasan mengenai persamaannya.

a. Normalisasi Data

Normalisasi data memastikan bahwa perhitungan jarak berdasarkan pengukuran studi kasusnya dan pengambilan data berdasarkan jarak terdekat dapat dilakukan dengan tepat. Persamaan dari normalisasi data ini adalah sebagai berikut:

$$X_{norm} = \frac{x_{data} - \min(x_{data})}{\max(x_{data}) - \min(x_{data})} \quad (4)$$

Keterangan:
X = Atribut data
X_{data} = Nilai data berdasarkan entri data
min(x) = Nilai absolut minimal dari keseluruhan entri data berdasarkan X_{data}
max(x) = Nilai absolut maksimal dari keseluruhan entri data berdasarkan X_{data}

b. Euclidean Distance

Dalam algoritma KNN, untuk menghitung jarak antara data, diterapkan metode perhitungan yang disebut Jarak Euclidean (Euclidean Distance). Jarak euclidean ini mengukur jarak antara data latih dan data uji dengan menggunakan persamaan sebagai berikut:

$$d(x,y) = \sqrt{\sum_{i=1}^{n}(xi - yi)^2} \quad (5)$$

Keterangan:
d = Jarak Euclidean
x = Data uji
y = Data latih
n = Dimensi data
i = Variabel data
xi = Data uji berdasarkan i
yi = Data latih berdasarkan i

Setelah euclidean distance ditemukan hasilnya, maka untuk mengetahui data yang paling dekat membutuhkan pemilihan/voting data berdasarkan jarak/panjang euclidean yang terkecil di pengurutan data dengan nilai K.

Copyright © 2023 **Lazuardi Firdaus**, Page 189
This Journal is licensed under a Creative Commons Attribution 4.0 International License

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 5, Nomor 2, Desember 2023
e-ISSN 2685-998X
DOI 10.30865/json.v5i2.6124

Data yang dipilih merupakan jumlah kelas terbanyak dari nilai K yang sudah didapatkan pada proses pengurutan data tersebut.

2.5 Evaluation

Evaluation (evaluasi data) adalah proses pengukuran dan penilaian kualitas data untuk menentukan sejauh mana data tersebut dapat diandalkan, akurat, lengkap, dan relevan. Evaluasi dilakukan secara mendalam bertujuan untuk mendapatkan model yang sesuai dengan tujuan [21]. Evaluasi data pada penelitian ini menggunakan confusion matrix untuk mendapatkan hasil klasifikasi yang dibagi berdasarkan data aktual dan data prediksinya. Format dari confusion matrix untuk evaluasi data pada RapidMiner adalah sebagai berikut.

Tabel 1. Format Confusion Matrix

		Actual	
		true No (negative)	**true Yes (positive)**
Prediction	prediction No (negative)	TN	FN
	prediction Yes (positive)	FP	TP

TP = Klasifikasi data prediksi positif yang benar dari contoh data aktual positif (True Positive)
FP = Klasifikasi data prediksi positif yang salah dari contoh data aktual negatif (False Positive)
TN= Klasifikasi data prediksi negatif yang benar dari contoh data aktual negatif (True Negative)
FN = Klasifikasi data prediksi negatif yang salah dari contoh data aktual positif (False Negative)

Pada aplikasi RapidMiner, parameter seperti akurasi model data akan digunakan untuk mengetahui kinerja dari algoritma klasifikasi yang digunakan untuk klasifikasi produk populer Adidas US. Akurasi model data terdapat di persamaan (6) yang dituliskan sebagai berikut.

$$\text{Accuracy} = \frac{\text{True Positive}+\text{True Negative}}{\text{True Positive}+\text{False Positive}+\text{False Negative}+\text{True Negative}} \quad (6)$$

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

Pengumpulan data awal pada penelitian ini dikumpulkan dengan mengambil dataset publik. Sumber data penelitian ini merupakan data sekunder yang diunggah oleh Crawl Feeds [22] dengan nama dataset yaitu "Adidas US retail products dataset". Atribut dataset yang dibutuhkan adalah atribut "name", "selling_price", "availability", "color", "category", "breadcrumbs", "average_rating", dan "reviews_count". Namun sebelum membuat atribut "popular_product" yang dibutuhkan untuk dijadikan sebagai label/target klasifikasi, maka proses seleksi data pada tahap data preparation (persiapan data) perlu dilakukan agar hubungan antar variabel dapat diketahui.

Dataset awal menampilkan data yang berisikan produk Adidas US yang memiliki 845 baris data dan 20 atribut data. Adapun data uji pada penelitian ini akan diambil secara langsung dari situs web Adidas America Inc. dengan rasio sebesar 20% dari data latih, dengan alamat situsnya yaitu https://www.adidas.com/us,yang mana data uji tersebut dikumpulkan menggunakan proses input data secara manual dengan aplikasi Microsoft Excel. Berikut adalah deskripsi atribut data awal yang didapatkan untuk penelitian ini.

Tabel 2. Deskripsi atribut awal pada dataset

Variabel	Keterangan	Tipe Data
index	Penomoran baris (record) dari dataset	Numerik
url	Alamat URL dengan tautan (link) setiap produk	Nominal
name	Nama produk	Nominal
sku	SKU (Stock Keeping Unit)	Nominal
selling_price	Harga jual produk	Numerik
original_price	Harga asli untuk jual produk	Numerik
currency	Mata uang untuk harga jual dan harga asli	Numerik
availability	Status ketersediaan barang atau produk	Nominal
color	Warna produk	Nominal
category	Kategori produk	Nominal
source	Sumber penyedia data produk yang disediakan	Nominal
source_website	Situs sumber dari mana data dikumpulkan.	Nominal
breadcrumbs	Kategori produk (navigasi) website Adidas US	Nominal
description	Penjelasan singkat produk Adidas US.	Nominal
brand	Merek produk.	Nominal

Copyright © 2023 **Lazuardi Firdaus**, Page 190
This Journal is licensed under a Creative Commons Attribution 4.0 International License

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 5, Nomor 2, Desember 2023
e-ISSN 2685-998X
DOI 10.30865/json.v5i2.6124

Variabel	Keterangan	Tipe Data
images	Gambar produk Adidas US.	Nominal
country	Negara asal/tujuan produk.	Nominal
language	Bahasa situs web.	Nominal
average_rating	Nilai rata-rata penilaian produk oleh pelanggan	Numerik
reviews_count	Jumlah ulasan produk oleh pelanggan	Numerik
crawled_at	Tanggal dan waktu saat data dikumpulkan	Timestamp

3.2 Data Pre-Processing

Data Pre-Processing atau pra-pemrosesan data merupakan tahapan dalam persiapan data untuk memproses data menjadi data yang berkualitas dan memiliki bentuk data yang dibutuhkan. Pada tahapan pra-pemrosesan data ini menggunakan model CRISP-DM seperti yang sudah tertera pada gambar 1 dan mengikuti tahapan penelitian yang dijelaskan pada gambar 2. Tahap Data Pre-Processing terdiri dari seleksi data, deskripsi data, pembersihan data, dan transformasi data, yang dilakukan untuk mempersiapkan data sebelum diolah lebih lanjut.

Proses seleksi data pada tahap ini mengambil atribut/variabel data dari tabel 2, yang mana variabel diseleksi adalah merupakan variabel yang benar-benar dibutuhkan seperti "name", "selling_price", "availability", "color", "category", "breadcrumbs", "average_rating", dan "reviews_count". Untuk deskripsi data menampilkan data yang sudah diseleksi agar dapat dilihat informasi mengenai data tersebut. Berikut ini adalah deskripsi data yang didapatkan setelah melalui proses seleksi data.

Tabel 3. Deskripsi dataset setelah diseleksi

Nama Atribut	Keterangan	Tipe Data
name	Nama produk	Nominal
selling_price	Harga jual produk	Numerik
availability	Status ketersediaan barang atau produk	Nominal
color	Warna produk	Nominal
category	Kategori produk	Nominal
breadcrumbs	Kategori produk berdasarkan navigasi website Adidas America, Inc.	Nominal
average_rating	Nilai rata-rata penilaian produk oleh pelanggan	Numerik
reviews_count	Jumlah ulasan produk oleh pelanggan	Numerik

Hasil seleksi data yang dijelaskan di tabel 3 merupakan data yang benar-benar dibutuhkan dalam proses modelling menggunakan algoritma klasifikasi dari Naive Bayes, Decision Tree, dan KNN. Setelah tahapan seleksi data dilakukan, maka pembersihan data dibutuhkan sesuai prosedur dari model CRISP-DM dan tahapan penelitian. Menurut Yunita dkk. [23], Pembersihan data (data cleaning) dilakukan untuk menghilangkan data yang tidak konsisten atau mengandung noise (kebisingan), yang merupakan salah satu tahapan dalam data mining. Proses pembersihan data ini sangat penting karena data yang tidak konsisten dapat menyebabkan hasil penambangan data yang tidak valid. Bentuk dataset untuk data latih setelah dilakukan pembersihan data memiliki data sebanyak 664 data atau sekitar 80% dari jumlah data sebelumnya. Jika data sudah dibersihkan, maka tahap berikutnya adalah melakukan transformasi data dari dataset yang digunakan.

Menurut Azhar dkk. [24], tujuan utama dari transformasi data adalah mengubah skala pengukuran data asli menjadi bentuk lain agar data tersebut sesuai dengan asumsi analisis varians yang mendasarinya. Dengan kata lain, proses transformasi digunakan untuk mengubah format data atau mengubah data sesuai dengan bentuk data yang diinginkan. Untuk melakukan hal tersebut, transformasi data diterapkan pada dataset Adidas US dengan tujuan mengubah data numerik menjadi nominal sehingga mempermudah proses pemodelan.

Parameter nilai kuartil statistik digunakan untuk membentuk data ordinal pada atribut "selling_price", "average_rating", dan "reviews_count". Untuk memudahkan dan meningkatkan efektivitas proses pemodelan dengan menggunakan algoritma KNN, data dengan tipe numerik diubah menjadi bentuk nominal yang bersifat ordinal. Dalam hal ini, terjadi sebuah pengecualian pada pengubahan format data dari dataset KNN karena algoritma tersebut memerlukan data numerik dalam proses normalisasi dan perhitungan euclidean distance-nya. Hasil akhir dari data pre-processing pada dataset tertera di gambar 3 dan 4 dengan tampilan datanya adalah sebagai berikut.

	name	availability	color	category	breadcrumbs	selling_price	average_rating	reviews_count	popular_product
0	Beach Shorts	InStock	Black	Clothing	Women/Clothing	Medium Price	Medium Rating	Medium Reviews Count	No
1	Five Ten Kestrel Lace Mountain Bike Shoes	InStock	Grey	Shoes	Women/Shoes	Very High Price	Very High Rating	Low Reviews Count	No
2	Mexico Away Jersey	InStock	White	Clothing	Kids/Clothing	Very High Price	Very High Rating	Medium Reviews Count	Yes
3	Five Ten Hiangle Pro Competition Climbing Shoes	InStock	Black	Shoes	Five Ten/Shoes	Very High Price	Low Rating	Low Reviews Count	No
4	Mesh Broken-Stripe Polo Shirt	InStock	Blue	Clothing	Men/Clothing	Very High Price	High Rating	Low Reviews Count	No

Gambar 3. Hasil data pre-processing untuk klasifikasi Naive Bayes dan Decision Tree

Copyright © 2023 **Lazuardi Firdaus**, Page 191
This Journal is licensed under a Creative Commons Attribution 4.0 International License

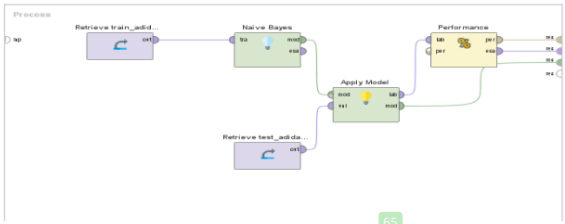
Jurnal Sistem Komputer dan Informatika (JSON)
Volume 5, Nomor 2, Desember 2023
e-ISSN 2685-998X
DOI 10.30865/json.v5i2.6124

	name	availability	color	category	breadcrumbs	selling_price	average_rating	reviews_count	popular_product
0	Beach Shorts	1	1	1	20	1	1	1	No
1	Five Ten Kestrel Lace Mountain Bike Shoes	1	7	2	21	3	3	0	No
2	Mexico Away Jersey	1	16	1	4	3	3	1	Yes
3	Five Ten Hiangle Pro Competition Climbing Shoes	1	1	2	2	3	0	0	No
4	Mesh Broken-Stripe Polo Shirt	1	2	1	7	3	2	0	No

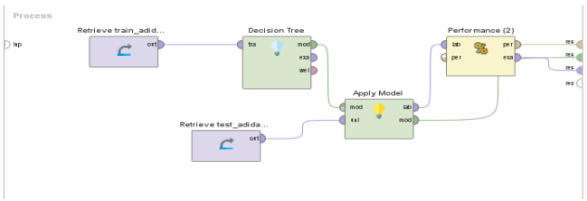
Gambar 4. Hasil data pre-processing untuk klasifikasi KNN

3.3 Uji Model Klasifikasi

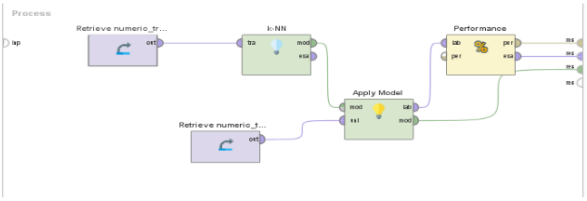
Pengujian model klasifikasi dilakukan untuk membuat model menggunakan model CRISP-DM (lihat gambar 1 dan gambar 2) dengan algoritma klasifikasi yang digunakan dalam klasifikasi produk populer Adidas US pada dataset yang sudah dibuat. Proses uji model ini menggunakan algoritma Naive Bayes, Decision Tree, dan KNN sebagai model klasifikasinya. Tahap uji model klasifikasi ini menggunakan aplikasi RapidMiner versi 10.1. Tahapan dari pembuatan model klasifikasi dapat dilihat pada gambar 5, 6, dan 7.



Gambar 5. Rancangan model klasifikasi Naive Bayes



Gambar 6. Rancangan model klasifikasi Decision Tree



Gambar 7. Rancangan model klasifikasi KNN

Model klasifikasi pada tahapan ini menggunakan data latih dan data uji dengan rasio data adalah data latih sebesar 80% dan data uji sebesar 20%. Rancangan model klasifikasi ini juga menggunakan operator Apply Model, yang berfungsi untuk menerapkan operator klasifikasi yaitu Naive Bayes, Decision Tree, dan KNN dari data latih ke data uji. Setelah model diterapkan, operator Performance digunakan untuk mengevaluasi pengujian model klasifikasi dengan menggunakan confusion matrix dan parameter lainnya.

3.3 Evaluasi Model Klasifikasi

Evaluasi model klasifikasi digunakan untuk mengetahui hasil klasifikasi data yang dihasilkan pada tahapan uji model (modelling) dengan metode CRISP-DM dan mengikuti tahapan penelitian. Ketika semua model klasifikasi

Copyright © 2023 **Lazuardi Firdaus**, Page 192
This Journal is licensed under a Creative Commons Attribution 4.0 International License

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 5, Nomor 2, Desember 2023
e-ISSN 2685-998X
DOI 10.30865/json.v5i2.6124

sudah diuji menggunakan rancangan model yang dibuat, maka operator Performance akan menghasilkan banyak data yang diprediksi berdasarkan data aktual pada confusion matrix dan nilai akurasi dari model tersebut. Confusion matrix (lihat tabel 1) terdiri dari True Positive (TP), False Positive (FP), True Negative (TN), dan False Negative (FN) yang digunakan untuk melihat hasil klasifikasi dan juga digunakan untuk mencari nilai akurasi dari model. Hasil klasifikasi dari ketiga model klasifikasi ini dengan confusion matrix adalah sebagai berikut.

Tabel 4. Confusion matrix model klasifikasi Naive Bayes

	true No	true Yes	class precision
pred. No	35	11	76.09%
pred. Yes	7	79	91.86%
class recall	83.33%	87.78%	

Tabel 5. Confusion matrix model klasifikasi Decision Tree

	true No	true Yes	class precision
pred. No	38	6	86.36%
pred. Yes	4	84	95.45%
class recall	90.48%	93.33%	

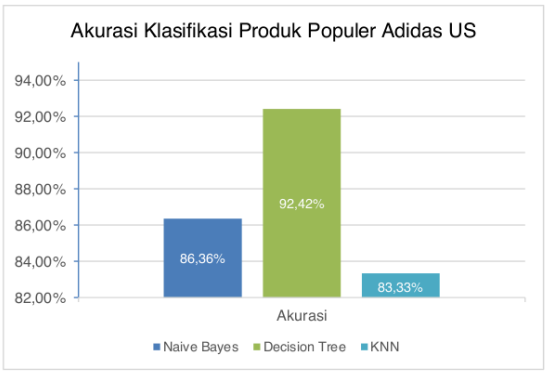
Tabel 6. Confusion matrix model klasifikasi KNN

	true No	true Yes	class precision
pred. No	27	7	79.41%
pred. Yes	15	83	84.69%
class recall	64.29%	92.22%	

Pada tabel 4, 5, dan 6, kesimpulan dari evaluasi data yang dapat diambil pada confusion matrix ini adalah sebagai berikut.

a. Hasil klasifikasi Naive Bayes menunjukkan bahwa terdapat 79 data klasifikasi positif yang benar dari contoh data aktual positif (true positive), 7 data klasifikasi positif yang salah dari contoh data aktual negatif (false positive), 35 data klasifikasi negatif yang benar dari contoh data aktual negatif (true negative), dan 11 data klasifikasi negatif yang salah dari contoh data aktual positif (false negative).
b. Sedangkan hasil klasifikasi dengan algoritma Decision Tree menunjukkan bahwa terdapat 84 data klasifikasi yang benar dari contoh data aktual positif (true positive), 4 data klasifikasi positif yang salah dari contoh data aktual negatif (false positive), 38 data klasifikasi negatif yang benar dari contoh data negatif (true negative), dan 6 data klasifikasi negatif yang salah dari contoh data aktual positif (false negative).
c. Hasil klasifikasi dengan algoritma KNN menunjukkan terdapat 83 data klasifikasi positif yang benar dari contoh data aktual positif (true positive), 15 data klasifikasi positif yang salah dari contoh data aktual negatif (false positive), 27 data klasifikasi negatif yang benar dari contoh data aktual negatif (true negative), dan 7 data klasifikasi negatif yang salah dari contoh data positif (false negative).

Ketika semua hasil klasifikasi data dari confusion matrix sudah diketahui, maka untuk langkah berikutnya adalah membandingkan accuracy dari algoritma Naive Bayes, Decision Tree, dan KNN. Berikut adalah hasil perbandingan model algoritma klasifikasi berdasarkan akurasinya.



Gambar 8. Akurasi klasifikasi produk populer Adidas US

Copyright © 2023 **Lazuardi Firdaus**, Page 193
This Journal is licensed under a Creative Commons Attribution 4.0 International License

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 5, Nomor 2, Desember 2023
e-ISSN 2685-998X
DOI 10.30865/json.v5i2.6124

Untuk hasil perbandingan algoritma klasifikasi ini menggunakan rasio sebesar 80% : 20% (data latih : data uji). Dari grafik di atas dapat dilihat bahwa performa dari algoritma Naive Bayes menghasilkan nilai akurasi sebesar 86,36%. Sedangkan pada performa dari algoritma Decision Tree menghasilkan nilai akurasi sebesar 92,42%. Untuk algoritma dari KNN menghasilkan nilai akurasi sebesar 83,33%. Dapat disimpulkan pada tahap evaluasi ini bahwa model Decision Tree memiliki performa yang lebih baik untuk klasifikasi data produk populer Adidas.

4. KESIMPULAN

Berdasarkan penelitian yang dilakukan, dapat disimpulkan bahwa penelitian ini menggunakan dataset yaitu "Adidas US retail products dataset“ yang digunakan untuk klasifikasi data produk populer Adidas US. Metode atau tahapan data mining yang digunakan pada tahapan penelitian adalah model CRISP-DM seperti yang tertera pada gambar 1, serta tahapan penelitian utama yang digunakan tertera pada gambar 2 yang mengacu pada model CRISP-DM tersebut. Pada tahap pemahaman bisnis, permasalahan yang dihadapi adalah kurangnya kategori produk populer untuk digunakan sebagai referensi dalam penjualan produk Adidas US dan dataset yang digunakan belum memiliki atribut yang dibutuhkan dan dipilah untuk dilakukan proses klasifikasi menggunakan algoritma klasifikasi Naive Bayes, Decision Tree, dan KNN. Tujuan utama dari penelitian ini adalah untuk menguji ketiga algoritma klasifikasi pada data produk populer Adidas US dan membandingkan hasilnya untuk mendapatkan klasifikasi produk populer yang akurat. Tujuan analisisnya adalah untuk menemukan hasil klasifikasi dengan tingkat akurasi yang tinggi dan membandingkan kinerja dari ketiga algoritma klasifikasi tersebut. Penelitian ini melakukan perbandingan algoritma klasifikasi dengan menggunakan tiga algoritma, yaitu Naive Bayes, Decision Tree, dan KNN. Dalam hasil klasifikasi data yang didapatkan pada tabel 4, 5, dan 6, Naive Bayes mengklasifikasikan 79 data aktual positif dengan benar sebagai positif (TP), 7 data aktual negatif salah terklasifikasi sebagai positif (FP), 35 data aktual negatif benar-benar terklasifikasi sebagai negatif (TN), dan 11 data aktual positif salah terklasifikasi sebagai negatif (FN). Decision Tree mengklasifikasikan 84 data aktual positif dengan benar sebagai positif (TP), 4 data aktual negatif salah terklasifikasi sebagai positif (FP), 38 data aktual negatif benar-benar terklasifikasi sebagai negatif (TN), dan 6 data aktual positif salah terklasifikasi sebagai negatif (FN). Sedangkan KNN mengklasifikasikan 83 data aktual positif dengan benar sebagai positif (TP), 15 data aktual negatif salah terklasifikasi sebagai positif (FP), 27 data aktual negatif benar-benar terklasifikasi sebagai negatif (TN), dan 7 data aktual positif salah terklasifikasi sebagai negatif (FN).Berdasarkan gambar 8 pada bagian hasil evaluasi model klasifikasi, untuk perbandingan algoritma klasifikasi pada evaluasi data menunjukkan bahwa penggunaan algoritma Decision Tree ini merupakan algoritma yang terbaik untuk klasifikasi produk populer Adidas US. Hasil perbandingan ini menggunakan rasio sebesar 80% : 20% (data latih : data uji). Algoritma Decision Tree mampu mengklasifikasikan produk populer Adidas US dengan nilai akurasi sebesar 92,42%. Sedangkan algoritma Naive Bayes mampu mengklasifikasikan produk populer Adidas US dengan nilai akurasi sebesar 92,42%. Untuk algoritma KNN dapat melakukan klasifikasi produk populernya dengan nilai akurasi sebesar 83,33%.

REFERENCES

[1] P. B. N. Setio, D. R. S. Saputro, dan B. Winarno, “Klasifikasi dengan Pohon Keputusan Berbasis Algoritme C4.5,” dalam PRISMA, Prosiding Seminar Nasional Matematika 3, Jurusan Matematika, FMIPA Universitas Negeri Semarang, Feb 2020, hlm. 64–71.

[2] C. Schröer, F. Kruse, dan J. M. Gómez, “A systematic literature review on applying CRISP-DM process model,” dalam Procedia Computer Science, Elsevier B.V., Feb 2021, hlm. 526–534. doi: 10.1016/j.procs.2021.01.199.

[3] M. Calero Peréz, M. B. Calisto, S. Bonilla, dan D. Riofrío, “Application of Machine Learning algorithms for the prediction of payment by agreement in a debt collection company with the CRISP-DM methodology,” dalam 3rd South American International Conference on Industrial Engineering and Operations Management, Asuncion, Paraguay: IEOM Society International, Jul 2022, hlm. 474–485. doi: 10.46254/SA03.20220112.

[4] J. W. Iskandar dan Y. Nataliani, “Perbandingan Naïve Bayes, SVM, dan k-NN untuk Analisis Sentimen Gadget Berbasis Aspek,” Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi), vol. 5, no. 6, hlm. 1120–1126, Des 2021, doi: 10.29207/resti.v5i6.3588.

[5] A. P. Permana, K. Ainiyah, dan K. F. H. Holle, “Analisis Perbandingan Algoritma Decision Tree, kNN, dan Naive Bayes untuk Prediksi Kesuksesan Start-up,” JISKA (Jurnal Informatika Sunan Kalijaga), vol. 6, no. 3, hlm. 178–188, Sep 2021, doi: 10.14421/jiska.2021.6.3.178-188.

[6] D. Tri Wiyanti dan Ainurrohmah, “Analisis Performa Algoritma Decision Tree, Naïve Bayes, K-Nearest Neighbor untuk Klasifikasi Zona Daerah Risiko Covid-19 di Indonesia,” Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK), vol. 10, no. 1, hlm. 115–122, 2023, doi: 10.25126/jtiik.2023105935.

[7] C. Anam dan H. B. Santoso, “Perbandingan Kinerja Algoritma C4.5 dan Naive Bayes untuk Klasifikasi Penerima Beasiswa,” Jurnal ENERGY, vol. 8, no. 1, hlm. 2088–4591, Mei 2018.

[8] Y. I. Kurniawan, “Perbandingan Algoritma Naive Bayes dan C.45 dalam Klasifikasi Data Mining,” Jurnal Teknologi Informasi dan Ilmu Komputer, vol. 5, no. 4, hlm. 455, Okt 2018, doi: 10.25126/jtiik.201854803.

[9] S. Raschka, “Model Evaluation, Model Selection, and Algorithm Selection in Machine Learning,” arXiv preprint, hlm. 1–49, Nov 2018.

Copyright © 2023 **Lazuardi Firdaus**, Page 194
This Journal is licensed under a Creative Commons Attribution 4.0 International License

[10] M. A. Hasanah, S. Soim, dan A. S. Handayani, "Implementasi CRISP-DM Model Menggunakan Metode Decision Tree dengan Algoritma CART untuk Prediksi Curah Hujan Berpotensi Banjir," Journal of Applied Informatics and Computing (JAIC), vol. 5, no. 2, hlm. 103–108, Okt 2021, doi: https://doi.org/10.30871/jaic.v5i2.3200.
[11] V. K. Singh, A. Singh, dan K. Joshi, "Fair CRISP-DM: Embedding Fairness in Machine Learning (ML) Development Life Cycle," dalam Proceedings of the 55th Hawaii International Conference on System Sciences, 2022, hlm. 1531–1540.
[12] S. Marzukhi, N. Awang, S. N. Alsagoff, dan H. Mohamed, "RapidMiner and Machine Learning Techniques for Classifying Aircraft Data," dalam Journal of Physics: Conference Series, IOP Publishing Ltd, Agu 2021, hlm. 1–8. doi: 10.1088/1742-6596/1997/1/012012.
[13] N. Hidayati, J. Suntoro, dan G. G. Setiaji, "Perbandingan Algoritma Klasifikasi untuk Prediksi Cacat Software dengan Pendekatan CRISP-DM," Jurnal Sains dan Informatika, vol. 7, no. 2, hlm. 117–126, Nov 2021, doi: 10.34128/jsi.v7i2.313.
[14] I. Alfitra Salam, K. Wahyudi Putra, S. Yuliatina, dan B. Nurina Sari, "Application of Naïve Bayes for Classification of Criteria for Potable Water with the CRISP-DM Method," Paradigma, vol. 25, no. 1, hlm. 1–5, Maret 2023, doi: 10.31294/paradigma.v25i1.1754.
[15] H. Annur, "Klasifikasi Masyarakat Miskin Menggunakan Metode Naive Bayes," ILKOM Jurnal Ilmiah, vol. 10, no. 2, hlm. 160–165, Agu 2018, doi: 10.33096/ilkom.v10i2.303.160-165.
[16] I. Sutoyo, "IMPLEMENTASI ALGORITMA DECISION TREE UNTUK KLASIFIKASI DATA PESERTA DIDIK," PILAR Nusa Mandiri, vol. 14, no. 2, hlm. 217–224, 2018, [Daring]. Tersedia pada: www.bsi.ac.id
[17] H. Hafizan dan A. N. Putri, "Penerapan Metode Klasifikasi Decision Tree Pada Status Gizi Balita Di Kabupaten Simalungun," KESATRIA: Jurnal Penerapan Sistem Informasi (Komputer & Manajemen), vol. 1, no. 2, hlm. 68–72, 2020.
[18] P. B. N. Setio, D. R. S. Saputro, dan B. Winarno, "PRISMA, Prosiding Seminar Nasional Matematika Klasifikasi dengan Pohon Keputusan Berbasis Algoritme C4.5," vol. 3, hlm. 64–71, 2020, [Daring]. Tersedia pada: https://journal.unnes.ac.id/sju/index.php/prisma/
[19] N. Djamsi, D. Rizki Chandranegara, dan Z. Sari, "Mendeteksi Ekspresi Wajah dengan Meninjau Iris Mata Menggunakan Metode Transformasi Hough dan K-Nearest Neighbor (KNN)," REPOSITOR, vol. 5, no. 1, hlm. 575–580, 2023.
[20] A. Nikmatul Kasanah, U. Pujianto, dan Muladi, "Penerapan Teknik SMOTE untuk Mengatasi Imbalance Class dalam Klasifikasi Objektivitas Berita Online Menggunakan Algoritma KNN," JURNAL RESTI (Rekayasa Sistem dan Teknologi Informasi), vol. 1, no. 3, hlm. 196–201, Agu 2019, Diakses: 11 Maret 2023. [Daring]. Tersedia pada: http://jurnal.iaii.or.id/index.php/RESTI/article/view/945
[21] A. Khumaidi, "DATA MINING FOR PREDICTING THE AMOUNT OF COFFEE PRODUCTION USING CRISP-DM METHOD," Jurnal Techno Nusa Mandiri, vol. 17, no. 1, hlm. 1–8, Feb 2020, doi: 10.33480/techno.v17i1.1240.
[22] Crawl Feeds, "Adidas US retail products dataset," data.world, 23 Oktober 2021. https://data.world/crawlfeeds/adidas-us-retail-products-dataset (diakses 19 April 2023).
[23] D. N. Yunita, A. H. S. Jones, dan D. Soyusiawaty, "Classification of Farmer's Eligibility as Recipients of Subsidized Fertilizer Assistance with C4.5 Algorithm," dalam IOP Conference Series: Materials Science and Engineering, Institute of Physics Publishing, Mar 2020. doi: 10.1088/1757-899X/771/1/012029.
[24] Y. Azhar, A. Khoiriyah Firdausy, dan P. J. Amelia, "Perbandingan Algoritma Klasifikasi Data Mining Untuk Prediksi Penyakit Stroke," SINTECH (Science and Information Technology), vol. 5, no. 2, hlm. 191–197, Okt 2022, doi: https://doi.org/10.31598.

Copyright © 2023 **Lazuardi Firdaus**, Page 195
This Journal is licensed under a Creative Commons Attribution 4.0 International License

cek klasifikasi adidas

ORIGINALITY REPORT

22%	18%	15%	9%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Yesi Pitaloka Anggriani, Alfis Arif, Febriansyah Febriansyah. "Implementasi Algoritma K-Means Clustering Dalam Menentukan Blok Tanaman Sawit Produktif Pada PT Arta Prigel", Jurnal Komtika (Komputasi dan Informatika), 2024 Publication	1%
2	Ilham Rizky Maziz Tomu, Marlon Stivo Noya Van Delsen, Norisca Lewaherilla. "KLASIFIKASI KETEPATAN WAKTU LULUS MAHASISWA JURUSAN MATEMATIKA UNIVERSITAS PATTIMURA MENGGUNAKAN METODE NAÏVE BAYES CLASSIFIER", Proximal: Jurnal Penelitian Matematika dan Pendidikan Matematika, 2024 Publication	1%
3	Submitted to Universitas Mercu Buana Student Paper	1%
4	digilib.uinsby.ac.id Internet Source	1%
5	Submitted to University of Southern Queensland Student Paper	1%
6	repository.uin-malang.ac.id Internet Source	1%
7	core.ac.uk Internet Source	1%
8	repository.its.ac.id	

	Internet Source	1 %
9	repository.uhamka.ac.id Internet Source	1 %
10	Submitted to STT PLN Student Paper	1 %
11	lab_adrk.ub.ac.id Internet Source	1 %
12	djournals.com Internet Source	<1 %
13	kc.umn.ac.id Internet Source	<1 %
14	jurnal.fikom.umi.ac.id Internet Source	<1 %
15	jurnal.publikasitecno.id Internet Source	<1 %
16	Submitted to Universitas PGRI Semarang Student Paper	<1 %
17	Submitted to Forum Perpustakaan Perguruan Tinggi Indonesia Jawa Timur Student Paper	<1 %
18	Yufis Azhar, Aidia Khoiriyah Firdausy, Putri Juli Amelia. "Perbandingan Algoritma Klasifikasi Data Mining Untuk Prediksi Penyakit Stroke", SINTECH (Science and Information Technology) Journal, 2022 Publication	<1 %
19	Submitted to University of Greenwich Student Paper	<1 %
20	journal.widyakarya.ac.id Internet Source	<1 %
21	jtiik.ub.ac.id Internet Source	

		<1 %
22	jurnal.ar-raniry.ac.id Internet Source	<1 %
23	repository.bsi.ac.id Internet Source	<1 %
24	publikasi.dinus.ac.id Internet Source	<1 %
25	Submitted to Universitas Brawijaya Student Paper	<1 %
26	jurnal.murnisadar.ac.id Internet Source	<1 %
27	www.teknobgt.com Internet Source	<1 %
28	idec.ft.uns.ac.id Internet Source	<1 %
29	scholar.archive.org Internet Source	<1 %
30	Dedy Hartama, Nanda Amalya. "Perbandingan Algoritma Decision Tree, ID3, dan Random Forest dalam Klasifikasi Faktor-Faktor yang Mempengaruhi Karier Mahasiswa Ilmu Komputer", Jurnal Indonesia : Manajemen Informatika dan Komunikasi, 2025 Publication	<1 %
31	Gelar Nurcahya, Arief Wibowo, Dwi Kristanto. "Perbandingan Kinerja Algoritma K-Means dan K-Medoids Dalam Klasterisasi Jumlah Tindak Pidana Kejahatan Berbasis Wilayah Kepolisian Daerah", SINTECH (Science and Information Technology) Journal, 2023 Publication	<1 %

32	Submitted to Padjadjaran University Student Paper	<1 %
33	Submitted to Politeknik Negeri Bandung Student Paper	<1 %
34	Submitted to Sriwijaya University Student Paper	<1 %
35	dqlab.id Internet Source	<1 %
36	jurusan.tik.pnj.ac.id Internet Source	<1 %
37	www.jurnal.stkippgritulungagung.ac.id Internet Source	<1 %
38	www.unaki.ac.id Internet Source	<1 %
39	Nurussakinah Nurussakinah, Muhammad Faisal. "Klasifikasi Penyakit Diabetes Menggunakan Algoritma Decision Tree", Jurnal Informatika, 2023 Publication	<1 %
40	ejournal.bsi.ac.id Internet Source	<1 %
41	jurnal.untan.ac.id Internet Source	<1 %
42	repository.upy.ac.id Internet Source	<1 %
43	www.researchgate.net Internet Source	<1 %
44	Submitted to Universitas Muhammadiyah Surakarta Student Paper	<1 %
45	Yohana Tri Widayati, Yani Prihati, Stephanus Widjaja. "ANALISIS DAN KOMPARASI	<1 %

ALGORITMA NAÏVE BAYES DAN C4.5 UNTUK KLASIFIKASI LOYALITAS PELANGGAN MNC PLAY KOTA SEMARANG", Jurnal Transformatika, 2021
Publication

46	ejurnal.teknokrat.ac.id Internet Source	<1 %
47	eprints.uad.ac.id Internet Source	<1 %
48	etheses.uin-malang.ac.id Internet Source	<1 %
49	id.scribd.com Internet Source	<1 %
50	jurnal.stmikasia.ac.id Internet Source	<1 %
51	ojs.cbn.ac.id Internet Source	<1 %
52	repository.ipb.ac.id Internet Source	<1 %
53	repository.ubharajaya.ac.id Internet Source	<1 %
54	www.coursehero.com Internet Source	<1 %
55	Dellatia Ayu Nur'Faradila, Lena Magdalena, Mesi Febima. "Penerapan Naïve Bayes Classifier Untuk Klasifikasi Postingan Berita Hoaks Di Instagram Cirebon Saber Hoaks", Media Jurnal Informatika, 2024 Publication	<1 %
56	Fitri Khoirunnisa, Salman Topiq. "ANALISIS SENTIMEN TERHDAP KEPERCAYAAN MASYARAKAT PADA PROSES PENEGAK HUKUM DI INDONESIA DENGAN	<1 %

MENGGUNAKAN ALGORITMA NAÏVE BAYES", Jurnal Informatika dan Teknik Elektro Terapan, 2024
Publication

57	Novanto Yudistira, Aldi Fianda Putra. "Algoritma Decision Tree Dan Smote Untuk Klasifikasi Serangan Jantung Miokarditis Yang Imbalance", Jurnal Litbang Edusaintech, 2021 Publication	<1 %
58	Resti Noor Fahmi, Mohamad Jajuli, Nina Sulistiyowati. "Analisis Pemetaan Tingkat Kriminalitas di Kabupaten Karawang menggunakan Algoritma K-Means", INTECOMS: Journal of Information Technology and Computer Science, 2021 Publication	<1 %
59	Yuliana Dilla Evitasari, Wawan Joko Pranoto, Naufal Adzmi Verdikha. "Evaluasi Support Vector Machine Dengan Optimasi Metode Genetic Algorithm Pada Klasifikasi Banjir Kota Samarinda", Jurnal Sains Komputer dan Teknologi Informasi, 2023 Publication	<1 %
60	docplayer.info Internet Source	<1 %
61	ejournal.uksw.edu Internet Source	<1 %
62	ejurnal.seminar-id.com Internet Source	<1 %
63	es.scribd.com Internet Source	<1 %
64	garuda.kemdikbud.go.id Internet Source	<1 %
65	ijcs.stmikindonesia.ac.id Internet Source	

		<1 %
66	jifosi.upnjatim.ac.id Internet Source	<1 %
67	journal.ibrahimy.ac.id Internet Source	<1 %
68	journal.uwks.ac.id Internet Source	<1 %
69	jutif.if.unsoed.ac.id Internet Source	<1 %
70	media.neliti.com Internet Source	<1 %
71	ojs.ninetyjournal.com Internet Source	<1 %
72	pdffox.com Internet Source	<1 %
73	repository.mercubuana.ac.id Internet Source	<1 %
74	repository.unja.ac.id Internet Source	<1 %
75	repository.unpas.ac.id Internet Source	<1 %
76	stmik-budidarma.ac.id Internet Source	<1 %
77	www.sciencegate.app Internet Source	<1 %
78	www.scribd.com Internet Source	<1 %
79	Herry Derajad Wijaya, Saruni Dwiasnati. "Implementasi Data Mining dengan Algoritma	<1 %

Naïve Bayes pada Penjualan Obat", Jurnal Informatika, 2020
Publication

80	Yuveinsiana Crismayella, Neva Satyahadewi, Hendra Perdana. "Algoritma Adaboost pada Metode Decision Tree untuk Klasifikasi Kelulusan Mahasiswa", Jambura Journal of Mathematics, 2023 Publication	<1%
81	jurnal.polsri.ac.id Internet Source	<1%

Exclude quotes	On	Exclude matches	Off
Exclude bibliography	On		