Indonesian Journal of Innovation Studies
Vol. 25 No. 3 (2024): July
DOI: 10.21070/ijins.v25i3.1156

Machine Learning Predicts Truck Breakdowns in Indonesia with 83% Accuracy: Machine Learning Memprediksi Kerusakan Truk di Indonesia dengan Akurasi 83%

Meisya Azzahra Rachman Universitas Muhammadiyah Sidoarjo
Tedjo Sukmono Universitas Muhammadiyah Sidoarjo

PT. Varia Usaha Beton, a cement product company, faces frequent breakdowns of mixer trucks, reducing reliability from the target 90% to 60%. This study aims to predict truck breakdowns using a machine learning model based on the K-NN algorithm within the CRISP-DM framework. Data from the company's maintenance records were cleaned and split into training and testing sets. With k=20, the model achieved 90% accuracy on training data and 83% on testing data. These results can help improve maintenance scheduling and resource planning, enhancing truck reliability. Future research should compare other algorithms and consider different programming environments.

Highlights:

1. High Accuracy: K-NN model achieved 90% training and 83% testing accuracy.
2. Maintenance Aid: Improves scheduling and resource planning for truck maintenance.
3. Future Research: Compare algorithms and explore different programming environments.

Keywords: Predictive Maintenance, Mixer Trucks, K-NN Algorithm, CRISP-DM, Machine Learning

Pendahuluan

Konstruksi adalah serangkaian pembangunan yang dirancang untuk memenuhi kebutuhan manusia [1]. Saat ini konstruksi yang banyak digunakan dan diminati adalah konstruksi beton karena memiliki banyak keunggulan dibandingkan dengan bahan lainnya, hal ini menjadikan penggunaan beton sebagai struktur konstruksi juga semakin meningkat [2]. Pemakaian alat berat dalam proyek-proyek konstruksi sangat membantu dalam pekerjaan suatu struktur bangunan, sehingga hasil yang diharapkan dapat tercapai dengan lebih mudah dalam waktu yang relatif lebih singkat [3].

Meningkatnya permintaan produksi beton dirasakan oleh PT.Varia Usaha Beton yang telah lama berpartisipasi sebagai penyedia jasa beton sejak tahun 1991, dan telah menyuplai produk ke berbagai proyek berskala besar di seluruh tanah air. Adapun alat penunjang yang digunakan dalam memenuhi kebutuhan pasar yang terus meningkat yaitu truk mixer, sebagai armada yang

menyuplai beton dari batching plant menuju lokasi proyek [4]. Namun, pada penggunaan truk mixer ini kerap mengalami breakdown sehingga perusahaan harus melakukan pergantian atau peremajaan mesin yang mengalami penurunan kualitas karena faktor masa ekonomis truk mixer yang berusia rata-rata lebih dari 15 tahun. Berdasarkan kebijakan perusahaan, target reliability truk mixer per bulannya sebesar 90% dari 10 unit yang ada, namun karena banyaknya kerusakan yang terjadi baik terjadwal maupun tidak terjadwal reliability yang diperoleh hanya sebesar 60%. Kerusakan yang umum terjadi seperti ban pecah, rem kurang pakem, tandon air bocor, talang air patah, dan lain-lain. Akibatnya selain berpengaruh pada persediaan suku cadang juga berakibat pada hal lainnya seperti: biaya-biaya yang berhubungan dengan persediaan suku cadang dan lamanya waktu tunggu pengiriman suku cadang.

Banyaknya breakdown yang terjadi, membuat prediksi kerusakan truk mixer di masa yang akan datang sangat dibutuhkan sebagai pencegahan dan pendeteksian risiko lebih awal. Dengan adanya pendeteksian sangat berguna bagi perusahaan dalam penyediaan stok barang, karena dengan prediksi yang dihasilkan dapat memberikan hasil terbaik agar risiko kesalahan yang ditimbulkan dapat ditekan seminimal mungkin [5]. Hasil dari suatu prediksi merupakan nilai kuantitatif terhadap subjek tertentu untuk jangka waktu tertentu. Prediksi selayaknya hanya sebagai masukan terhadap suatu perencanaan dan dapat berbeda dari rencana yang diperkirakan [6]. Identifikasi kerusakan truk mixer berdasarkan bagian kerusakan, tingkat risiko, frekuensi melalui proses klasifikasi. Klasifikasi adalah suatu proses yang digunakan untuk mendeskripsikan dan memetakan data ke dalam kelompok atau kelas yang telah ditentukan serta dapat meramalkan kecenderungan data pada masa depan [7]. Ada banyak teknik klasifikasi yang dapat digunakan salah satunya yaitu K-Nearest Neighbor (K-NN). Umumnya algoritma K-NN digunakan untuk mengklasifikasi objek berdasarkan pada data pembelajaran yang memiliki nilai selisih kecil dan jarak tetangga terdekat dengan objek [8]. Dipilihnya algoritma ini karena kemampuannya dalam menentukan kelompok data dengan menghitung jarak tetangga terdekat. Data yang diolah akan menjadi sebuah pengetahuan dan informasi yang berguna bagi banyak orang dalam melakukan prediksi dan estimasi sesuatu di masa yang akan datang. Oleh karena itu perlu dilakukannya data mining dalam proses mencari pola dari suatu data yang besar [9]. Algoritma K-NN mampu memproses data banyak dan mampu menghasilkan akurasi yang sangat baik dalam perbandingan kedekatan kasus baru dan kasus lama dan perlu menentukan nilai k [10]. Namun penentuan nilai k adalah kekurangan dari K-NN, penentuan nilai k harus dilakukan uji coba hingga menghasilkan tingkat akurasi yang optimal [11].

Penelitian terkait mengenai kegiatan prediksi dengan menggunakan algoritma K-Nearest Neighbor (K-NN) yaitu deteksi kerusakan pada jalan raya. Sampel data yang diuji terdiri dari 10 citra dengan label rusak dan tidak rusak. Hasil yang diperoleh menyatakan bahwa hasil yang sesuai sebanyak 8 dan hasil yang tidak sesuai sebanyak 2, maka persentase akurasi yang diperoleh adalah 90%. Maka, dapat dikatakan bahwa sistem yang dirancang dengan metode klasifikasi K-NN dapat digunakan dalam mendeteksi adanya kerusakan pada permukaan aspal jalan [12].

Penelitian kedua yang dilakukan untuk prediksi banjir dengan membandingkan nilai accuracy dan error antara algoritma K-NN dan Naive Bayes. Pemilihan data cuaca meliputi variabel target dan prediktor. Jumlah data yang akan diolah sejumlah 624 data. Data latih sebanyak 561 dan data uji sebanyak 63 data. Dengan k = 5, diperoleh nilai performansi setiap algoritma: K-NN memperoleh akurasi sebesar 88,94% dan error 11,06%. Sedangkan Naive Bayes memperoleh akurasi sebesar 74,36% dan error 25,64%. Dapat ditarik kesimpulan K-NN adalah algoritma terbaik guna memperkirakan tingkat kelulusan yang diinginkan [13].

Penelitian ketiga yang dilakukan untuk prediksi perubahan suhu yang ada di Indonesia dan mengukur tingkat akurasi dari prediksi tersebut. Data yang digunakan adalah data TEMP, karena banyak berisikan record yang memiliki atribut yang berbeda-beda, maka dataset akan dikelompokkan dalam tiga kelompok yaitu: suhu < = 70 Fahrenheit, suhu > 70 dengan < 80 Fahrenheit, dan suhu > = 80 Fahrenheit. Pada tahapan data, mulanya terdapat 28 atribut terpangkas menjadi 10 atribut yang akan dilanjutkan dalam proses pemodelan. Split data dilakukan

sebesar 80:20 dari 3623 data. Dimana data latih sebanyak 2.898 dan data uji sebanyak 725 data. Berdasarkan hasil yang diperoleh nilai akurasi sebesar 89% atau nilai tersebut masih terhitung cukup besar karena mendekati 100% [14].

Metode

Metode penelitian yang akan dilakukan menggunakan metode penelitian deskriptif analisis yang berfungsi dalam mendeskripsikan dan memberikan sebuah gambaran pada objek penelitian melalui pengumpulan data. Berikut alur penelitian yang dilakukan dalam prediksi kerusakan truk mixer pada PT. Varia Usaha Beton yang dapat dilihat pada gambar 1.

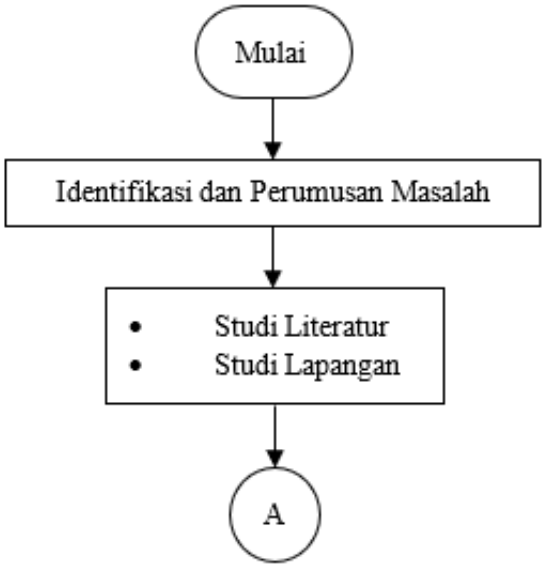


Figure 1. *Diagram Alir Penelitian*

Indonesian Journal of Innovation Studies
Vol. 25 No. 3 (2024): July
DOI: 10.21070/ijins.v25i3.1156

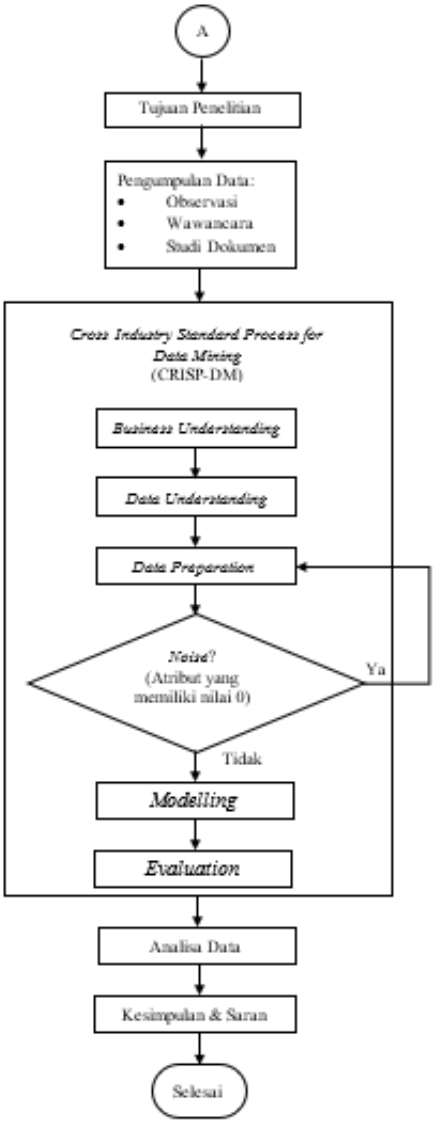


Figure 2. *Diagram Alir Penelitian*

A.Business Understanding

Pada tahapan ini dilakukan penetapan masalah yang akan diteliti yaitu mengenai kerusakan truk mixer pada PT. Varia Usaha Beton di masa mendatang. Kerusakan sangat penting untuk segera diselesaikan, hal yang harus diperhatikan adalah tanda-tanda yang muncul dapat dipahami dan diperhatikan sebelum terjadi kerusakan yang lebih parah. Hampir 99 persen kerusakan yang terjadi akan ditandai dengan munculnya tanda atau indikasi tertentu yang menunjukkan bahwa akan terjadinya kerusakan [15]. Tahapan selanjutnya yang harus dilakukan adalah mencari data kerusakan pada bengkel pemeliharaan PT. Varia Usaha Beton plant BSP Lingkar Timur.

B. Data Understanding

Selanjutnya proses mengumpulkan, mengidentifikasi, serta memahami data yang dimiliki. Data yang digunakan adalah data mengenai kerusakan truk mixer yang diperoleh dari data preventive maintenance harian PT. Varia Usaha Beton plant BSP Lingkar Timur.

C. Data Preparation

Indonesian Journal of Innovation Studies
Vol. 25 No. 3 (2024): July
DOI: 10.21070/ijins.v25i3.1156

Tahapan ini dilakukan untuk membangun dataset final dan penyempurnaan data yang akan dilanjutkan ke proses berikutnya. Sebelum membangun dataset final dilakukannya pembersihan data dengan bantuan Google Colaboratory menggunakan library Pandas pada bahasa pemrograman Phyton. Remove duplicates adalah suatu proses pembersihan data dengan membuang data-data yang sama atau diambil secara berulang pada saat proses scrapping [16]. Pembersihan data berguna untuk membuang data yang tidak memiliki nilai (null) karena akan mempengaruhi performansi dari sistem data mining karena berkurangnya jumlah dan kompleksitasnya [17].

D. Modelling

Pada tahapan ini akan dilakukan pemodelan terhadap dataset final. Dataset akan dibagi menjadi data latih dan data uji. Data training atau data latih adalah data yang digunakan pada modelling dalam proses klasifikasi. Data testing atau data uji adalah data yang digunakan pada proses prediksi dalam proses klasifikasi [18]. Pengolahan data dilakukan dengan proses data mining menggunakan CRISP-DM serta algoritma K-NN sebagai perhitungannya. Pada split data akan dibagi menjadi empat variasi yaitu 60:40, 70:30, 80:20, dan 90:10. Pembagian data latih dan data uji pada dataset dilakukan secara random dengan bantuan Phyton. Dari pembagian data yang dilakukan secara random ini membuat setiap data dalam sebuah kelas memiliki kemungkinan untuk menjadi data latih dan data uji [19]. Maka dari itu, proses pengacakan dataset menggunakan random state = 42, yang berguna agar hasil pembagian data tetap sama sewaktu akan di running kembali [20]. Nilai 42 yang ditetapkan adalah angka random. Nilai k yang menghasilkan akurasi tertinggi dapat dilihat pada gambar 2. Penentuan nilai k dibantu dengan menggunakan Google Colaboratory, maka dipilih nilai k=20 yang akan digunakan dalam mengambil jarak tetangga terdekat. Adapun perhitungan jarak tetangga dilakukan melalui perhitungan data latih dengan euclidean distance, untuk memperoleh jarak terdekat [21].

$$d\left(x_{1,},x_2\right)=\sqrt{\sum_{i=0}^{n}\left(x_{1i}\text{-}x_{2i}\right)^2} \quad (1)$$

Figure 3.

Dalam implementasinya pada model machine learning, rumus perhitungan euclidean distance dapat disesuaikan dengan banyaknya independent variabel pada dataset yang digunakan sebagai data latih. Jika terdapat data latih sebanyak dua atau lebih independent variabel begitu juga dimensi yang dihitung bertambah, maka formulanya seperti berikut [22]:

$$d=\sqrt{\sum_{i=1}^{n}\left(x_{1i}\text{-}\,x_{2i}\right)^2+\ \left(y_{1i}-y_{2i}\right)^2+\ +\left(z_{1i}-z_{2i}\right)^2+\cdots} \quad (2)$$

Figure 4.

Indonesian Journal of Innovation Studies
Vol. 25 No. 3 (2024): July
DOI: 10.21070/ijins.v25i3.1156

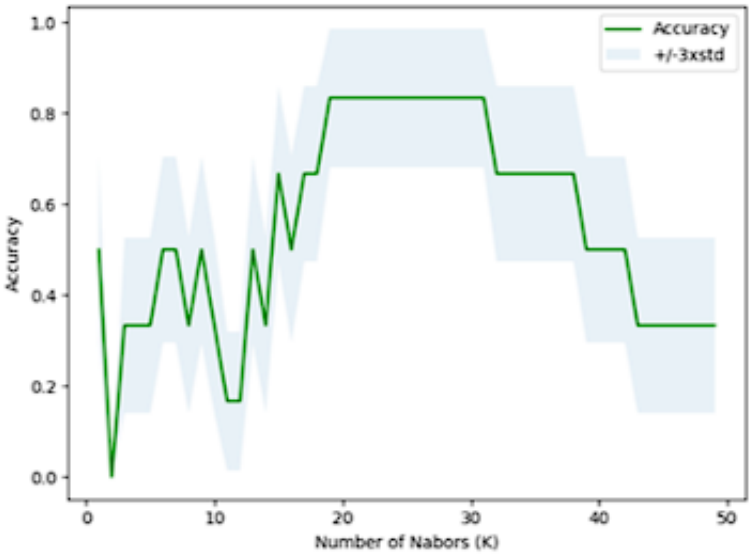


Figure 5. *Grafik Akurasi Setiap Parameter k*

E.Evaluation

Pada tahapan ini dilakukan evaluasi terhadap hasil perhitungan algoritma K-NN. Evaluasi yang dilakukan apakah sudah mencapai tujuan yang ditetapkan pada business understanding [23]. Evaluasi yang dilakukan dilihat dari nilai akurasi yang dihasilkan dengan menggunakan confusion matrix.

$$\text{Akurasi} = \frac{\text{Jumlah data terprediksi benar}}{\text{Jumlah seluruh data}} \times 100\% \quad (3)$$

Figure 6.

Hasil dan Pembahasan

A.Business Understanding

Pada tahapan business understanding, masalah yang akan diangkat adalah seberapa besar akurasi dan error yang diperoleh dengan menggunakan algoritma K-NN. Data yang akan digunakan adalah data kerusakan yang didapat dari data laporan preventive maintenance harian periode Januari 2018-Desember 2022.

a. Menilai Situasi

Data yang digunakan berguna sebagai informasi masih bersifat mentah atau belum diolah. Sehingga dilakukan pengolahan data agar siap digunakan dalam proses klasifikasi dari tahap awal hingga akhir. Kemudian, tidak adanya pendeteksian kerusakan guna menunjang kegiatan preventive maintenance yang telah dilakukan oleh PT. Varia Usaha Beton.

b. Tujuan Data Mining

Tujuan dari data mining adalah untuk mengeksplorasi perolehan informasi data kerusakan truk

mixer. Kemudian memprediksi hasil dengan menggunakan algoritma machine learning berupa K-NN dengan nilai k optimal.

B. Data Understanding

Pada tahapan ini dilakukan pengumpulan data mengenai kerusakan truk mixer dari periode Januari 2018-Desember 2022. Dataset mempunyai empat kriteria yaitu jenis kerusakan, tingkat risiko, frekuensi kerusakan dan bagian kerusakan. Pengujian dilakukan dengan menggunakan empat variasi dataset yang berbeda dengan jumlah data sebesar 60 records data kerusakan. Pembobotan pada tingkat risiko dilakukan oleh teknisi dan kepala bengkel perawatan truk mixer di PT. Varia Usaha Beton plant BSP Lingkar Timur. Variabel jenis kerusakan dapat dilihat pada tabel 1. Variabel penilaian tingkat risiko dapat dilihat pada tabel 2. Sedangkan variabel bagian kerusakan dapat dilihat pada tabel 3.

No	Kerusakan	Bagian Kerusakan
1	Handrem	Chasis unit
2	Tempat aki	Chasis unit
3	Slebor	Chasis unit
4	Pedal rem	Chasis unit
5	Rem	Chasis unit
...	...	...
56	Tensioner fan belt	Mesin
57	Handle pintu kabin	Kabin
58	Pintu kabin	Kabin
59	Gagang spion	Kabin
60	Kaca kabin	Kabin

Table 1. *Variabel Jenis Kerusakan*

Tingkat Risiko	Konversi Nilai
Ringan	1
Sedang	2
Tinggi	3

Table 2. *Variabel Penilaian Tingkat Risiko*

Bagian Kerusakan
Chasis Unit
Chasis Molen
Listrik
Mesin
Kabin

Table 3. *Variabel Bagian Kerusakan*

Dataset yang akan digunakan akan dibagi menjadi atribut dan label, dimana kolom jenis kerusakan digunakan sebagai id setiap records dan kolom bagian kerusakan sebagai label. Atribut yang digunakan yaitu tingkat risiko dan frekuensi kerusakan. Data kerusakan truk mixer selama 5 tahun dapat dilihat pada tabel 4.

Jenis Kerusakan	Tingkat Risiko	Frekuensi Kerusakan (2018)	Frekuensi Kerusakan (2019)	Frekuensi Kerusakan (2020)	Frekuensi Kerusakan (2021)	Frekuensi Kerusakan (2022)	Bagian Kerusakan
1	1	24	21	23	20	21	Chasis unit
2	1	5	1	4	3	2	Chasis unit
3	1	5	7	5	5	6	Chasis unit
4	2	10	12	9	9	11	Chasis unit
5	2	27	24	30	28	26	Chasis unit

...	...	...	...	...	...	...	...
56	2	2	3	2	1	4	Mesin
57	2	3	2	1	3	1	Kabin
58	2	3	4	3	5	4	Kabin
59	2	3	1	1	2	1	Kabin
60	2	5	6	5	6	6	Kabin

Table 4. *Data Kerusakan Truk Mixer*

C. Data Preparation

Berdasarkan tabel 4, data kerusakan truk mixer diubah ke dalam format sederhana agar dapat terbaca oleh sistem dengan format csv. Dataset final berupa empat atribut diantaranya yaitu: jenis kerusakan, tingkat risiko, frekuensi kerusakan selama 5 tahun (tahun 2018-2022), dan bagian kerusakan. Pada proses remove duplicates hasil yang diperoleh, tidak adanya missing value pada dataset maka jumlah data awal dengan jumlah data pada tahapan ini sama. Dataset final tersaji pada tabel 5.

Jenis Kerusakan	Tingkat Risiko	Frekuensi Kerusakan	Bagian Kerusakan
1	1	109	Chasis unit
2	1	15	Chasis unit
3	1	28	Chasis unit
4	2	51	Chasis unit
5	2	135	Chasis unit
...	...	...	...
...	...	...	...
56	2	12	Mesin
57	2	10	Kabin
58	2	19	Kabin
59	2	8	Kabin
60	2	28	Kabin

Table 5. *Dataset Final*

D. Modelling

Pada tahap modelling ini dilakukan pengukuran performa klasifikasi dengan menggunakan algoritma K-NN pada dataset kerusakan truk mixer. Pembagian data pada tahap modelling menggunakan data latih dan data uji 90% : 10%. Perbandingan tersebut menjadikan jumlah data latih sebesar 54 data dan data uji sebesar 6 data yang akan digunakan dalam proses pemodelan. Adapun nilai k yang digunakan k=20. Data yang digunakan akan di acak agar memiliki kesempatan menjadi data latih dan data uji. Data latih dan data uji yang digunakan pada pembagian data 90% data latih dan 10% data uji dapat dilihat pada tabel 6 dan tabel 7.

No	Jenis Kerusakan	Tingkat Risiko	Frekuensi Kerusakan	Bagian Kerusakan
1	34	2	44	Chasis molen
2	49	2	87	Mesin
3	13	2	76	Chasis unit
4	58	2	19	Kabin
5	47	2	32	Mesin
...	...	...	...	...
...	...	...	...	...
50	43	2	8	Mesin

51	15	2	123	Chasis unit
52	29	2	83	Chasis molen
53	52	2	15	Mesin
54	39	2	40	Listrik

Table 6. *Data Latih*

No	Jenis Kerusakan	Tingkat Risiko	Frekuensi Kerusakan	Bagian Kerusakan
1	1	1	109	Chasis unit
2	6	2	102	Chasis unit
3	37	2	60	Chasis molen
4	46	2	37	Mesin
5	14	1	55	Chasis unit
6	55	2	27	Mesin

Table 7. *Data Uji*

Selanjut perhitungan jarak menggunakan euclidean distance, perhitungan manual berikut adalah contoh perhitungan pada data latih 90% dan data uji 10%. Nilai yang diambil yaitu baris ke satu dari data latih dan data uji. Berikut perhitungan manual euclidean distance:

$$d = \sqrt{\sum_{i=1}^{n} (x_{1i} - x_{2i})^2 + (y_{1i} - y_{2i})^2}$$
$$d = \sqrt{(2-1)^2 + (44-109)^2}$$
$$d = 65{,}0077$$

Figure 7.

Berikut adalah hasil perhitungan euclidean yang dilakukan secara menyeluruh agar memperoleh nilai jarak antar data. Berikut adalah hasil jarak tetangga terdekat dari data lama dan data baru yang terlihat pada tabel 8.

No	Jenis Kerusakan	Tingkat Risiko	Frekuensi Kerusakan	Bagian Kerusakan
1	34	2	44	Chasis molen
2	49	2	87	Mesin
3	13	2	76	Chasis unit
4	58	2	19	Kabin
5	47	2	32	Mesin
...	...	...	...	...
...	...	...	...	...
54	39	2	40	Listrik

Table 8. *Hasil Jarak Tetangga Terdekat*

Selanjutnya adalah mengurutkan jarak data dari yang terkecil hingga terbesar, hasil jarak data yang telah diurutkan dapat dilihat pada tabel 9.

Jenis Kerusakan	Tingkat Risiko	Frekuensi Kerusakan	Bagian Kerusakan	Euclidean Distance
31	2	110	Chasis molen	1,4142
44	2	113	Mesin	4,1231
28	2	99	Chasis molen	10,0499

Indonesian Journal of Innovation Studies
Vol. 25 No. 3 (2024): July
DOI: 10.21070/ijins.v25i3.1156

32	1	122	Chasis molen	13,0000
15	2	123	Chasis unit	14,0357
...	...	...	...	...
...	...	...	...	...
17	2	256	Chasis unit	147,0034

Table 9. *Hasil Jarak Data Setelah Diurutkan*

Selanjutnya adalah pengujian dengan menggunakan 90% data latih dan 10% data uji, untuk melihat hasil prediksi kerusakan dengan perbandingan kelas yang ada. Hasil perbandingan dapat dilihat pada tabel 10.

No	Jenis Kerusakan	Tingkat Risiko	Frekuensi Kerusakan	Bagian Kerusakan (Asli)	Bagian Kerusakan (Prediksi)	Hasil Pengujian
1	1	1	109	Chasis unit	Chasis unit	Benar
2	6	2	102	Chasis unit	Chasis unit	Benar
3	37	2	60	Chasis molen	Chasis unit	Salah
4	46	2	37	Mesin	Mesin	Benar
5	14	1	55	Chasis unit	Chasis unit	Benar
6	55	2	27	Mesin	Mesin	Benar

Table 10. *Hasil Uji Coba*

E. Evaluation

Dari hasil pengujian yang dilakukan pada tahap modelling selanjutnya akan dilakukan evaluasi hasil untuk mengukur kinerja algoritma K-NN dengan melihat tingkat akurasi dengan memperhatikan confusion matrix. Confusion matrix yang pada dasarnya digunakan untuk melakukan perhitungan akurasi pada konsep data mining. Keakuratan hasil klasifikasi dapat diukur dengan menggunakan confusion matrix [24]. Berikut hasil evaluasi model yang diambil dari empat variasi dataset yang digunakan dalam penelitian, dapat dilihat pada beberapa gambar dan tabel berikut.

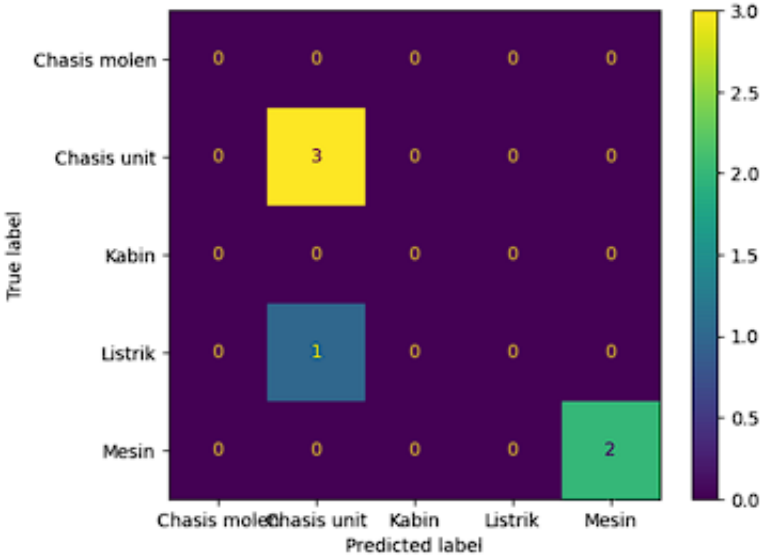


Figure 8. *Confusion Matrix Pada Data Latih 90% dan Data Uji 10%*

Indonesian Journal of Innovation Studies
Vol. 25 No. 3 (2024): July
DOI: 10.21070/ijins.v25i3.1156

Berdasarkan pada gambar 3, perhitungan nilai akurasi dari confusion matrix menggunakan persamaan sebagai berikut:

$$\text{Akurasi} = \frac{\text{Jumlah data terprediksi benar}}{\text{Jumlah seluruh data}} \times 100\%$$

$$\text{Akurasi} = \frac{3+2}{6} \times 100\%$$

$$\text{Akurasi} = 0{,}83$$

Figure 9.

Jenis Kerusakan	Tingkat Risiko	Frekuensi Kerusakan	Bagian Kerusakan (Asli)	Bagian Kerusakan (Prediksi)	Hasil Pengujian
1	1	109	Chasis unit	Chasis unit	Benar
6	2	102	Chasis unit	Chasis unit	Benar
37	2	60	Chasis molen	Chasis unit	Salah
46	2	37	Mesin	Mesin	Benar
14	1	55	Chasis unit	Chasis unit	Benar
55	2	27	Mesin	Mesin	Benar

Table 11. *Hasil Uji Coba Pada Data Latih 90% dan Data Uji 10%*

Pada tabel 11, dapat disimpulkan pada pembagian data 90% data latih dan 10% data uji, diperoleh kelas terbanyak pada hasil prediksi yaitu chasis unit. Hasil kelas terbanyak ini menjadi hasil kerusakan yang diprediksi akan terjadi di masa mendatang dengan akurasi sebesar 83%, hasil akurasi yang diperoleh dibuktikan dengan perbedaan dari hasil data prediksi yang tidak sesuai dengan data asli.

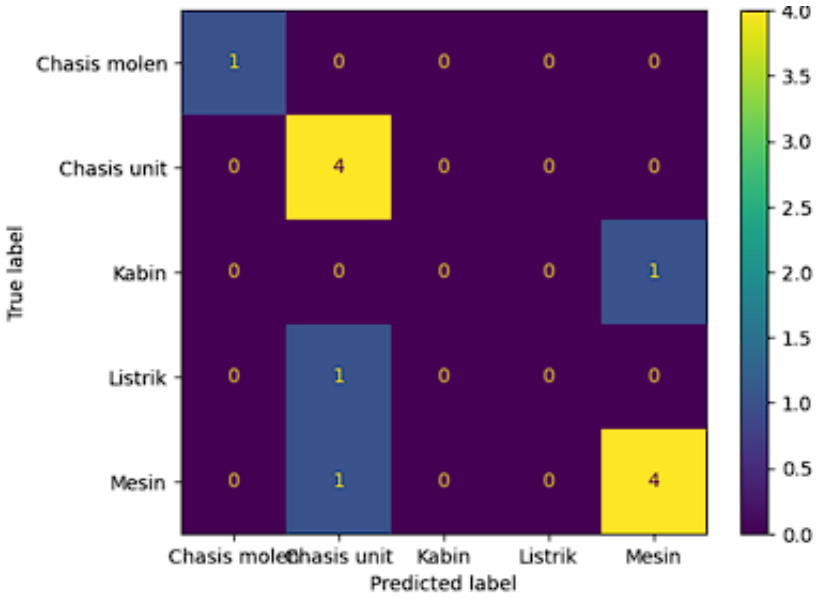


Figure 10. *Confusion Matrix Pada Data Latih 80% dan Data Uji 20%*

Berdasarkan pada gambar 4, perhitungan nilai akurasi dari confusion matrix menggunakan persamaan sebagai berikut:

Indonesian Journal of Innovation Studies
Vol. 25 No. 3 (2024): July
DOI: 10.21070/ijins.v25i3.1156

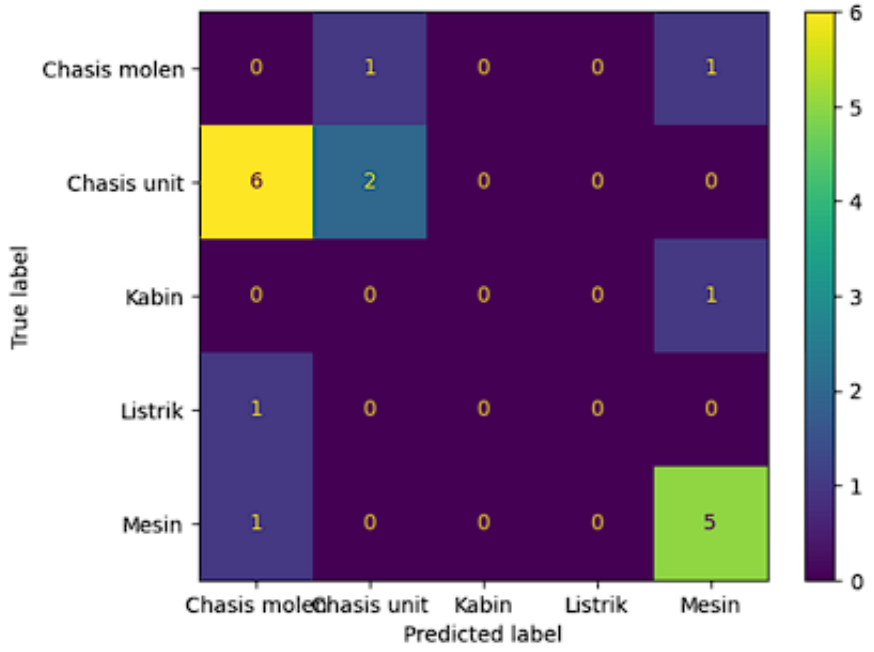
$$\text{Akurasi} = \frac{\text{Jumlah data terprediksi benar}}{\text{Jumlah seluruh data}} \times 100\%$$
$$\text{Akurasi} = \frac{4+1+4}{12} \times 100\%$$
$$\text{Akurasi} = 0{,}75$$

Figure 11.

Jenis Kerusakan	Tingkat Risiko	Frekuensi Kerusakan	Bagian Kerusakan (Asli)	Bagian Kerusakan (Prediksi)	Hasil Pengujian
1	1	109	Chasis unit	Chasis unit	Benar
6	2	102	Chasis unit	Chasis unit	Benar
37	2	60	Listrik	Chasis unit	Salah
46	2	37	Mesin	Mesin	Benar
14	1	55	Chasis unit	Chasis unit	Benar
55	2	27	Mesin	Mesin	Benar
34	2	44	Chasis molen	Chasis molen	Benar
49	2	87	Mesin	Chasis molen	Salah
13	2	76	Chasis unit	Chasis unit	Benar
58	2	19	Kabin	Mesin	Salah
47	2	32	Mesin	Mesin	Benar
51	2	13	Mesin	Mesin	Benar

Table 12. *Confusion Matrix Pada Data Latih 80% dan Data Uji 20%*

Pada tabel 12, dapat disimpulkan pada pembagian data 80% data latih dan 20% data uji, diperoleh kelas terbanyak pada hasil prediksi yaitu chasis unit dan mesin. Hasil kelas terbanyak ini menjadi hasil kerusakan yang diprediksi akan terjadi di masa mendatang dengan akurasi sebesar 75%, hasil akurasi yang diperoleh dibuktikan dengan perbedaan dari hasil data prediksi yang tidak sesuai dengan data asli.



Indonesian Journal of Innovation Studies
Vol. 25 No. 3 (2024): July
DOI: 10.21070/ijins.v25i3.1156

Figure 12. *Confusion Matrix Pada Data Latih 70% dan Data Uji 30%*

Berdasarkan pada gambar 5, perhitungan nilai akurasi dari confusion matrix menggunakan persamaan sebagai berikut:

$$\text{Akurasi} = \frac{\text{Jumlah data terprediksi benar}}{\text{Jumlah seluruh data}} \times 100\%$$

$$\text{Akurasi} = \frac{2+5}{18} \times 100\%$$

$$\text{Akurasi} = 0{,}39$$

Figure 13.

Jenis Kerusakan	Tingkat Risiko	Frekuensi Kerusakan	Bagian Kerusakan (Asli)	Bagian Kerusakan (Prediksi)	Hasil Pengujian
1	1	109	Chasis unit	Chasis molen	Salah
6	2	102	Chasis unit	Chasis molen	Salah
37	2	60	Listrik	Chasis molen	Salah
46	2	37	Mesin	Mesin	Benar
14	1	55	Chasis unit	Chasis molen	Salah
55	2	27	Mesin	Mesin	Benar
34	2	44	Chasis molen	Chasis molen	Benar
49	2	87	Mesin	Chasis molen	Salah
13	2	76	Chasis unit	Chasis molen	Salah
58	2	19	Kabin	Mesin	Salah
47	2	32	Mesin	Mesin	Benar
51	2	13	Mesin	Mesin	Benar
32	1	122	Chasis molen	Chasis molen	Benar
4	2	51	Chasis unit	Chasis molen	Salah
53	2	35	Mesin	Mesin	Benar
18	2	80	Chasis unit	Chasis molen	Salah
9	2	91	Chasis unit	Chasis molen	Salah
7	2	65	Chasis unit	Chasis molen	Salah

Table 13. *Confusion Matrix Pada Data Latih 70% dan Data Uji 30%*

Pada tabel 13, dapat disimpulkan pada pembagian data 70% data latih dan 30% data uji, diperoleh kelas terbanyak pada hasil prediksi yaitu chasis molen. Hasil kelas terbanyak ini menjadi hasil kerusakan yang diprediksi akan terjadi di masa mendatang dengan akurasi sebesar 39%, hasil akurasi yang diperoleh dibuktikan dengan perbedaan dari hasil data prediksi yang tidak sesuai dengan data asli.

Indonesian Journal of Innovation Studies
Vol. 25 No. 3 (2024): July
DOI: 10.21070/ijins.v25i3.1156

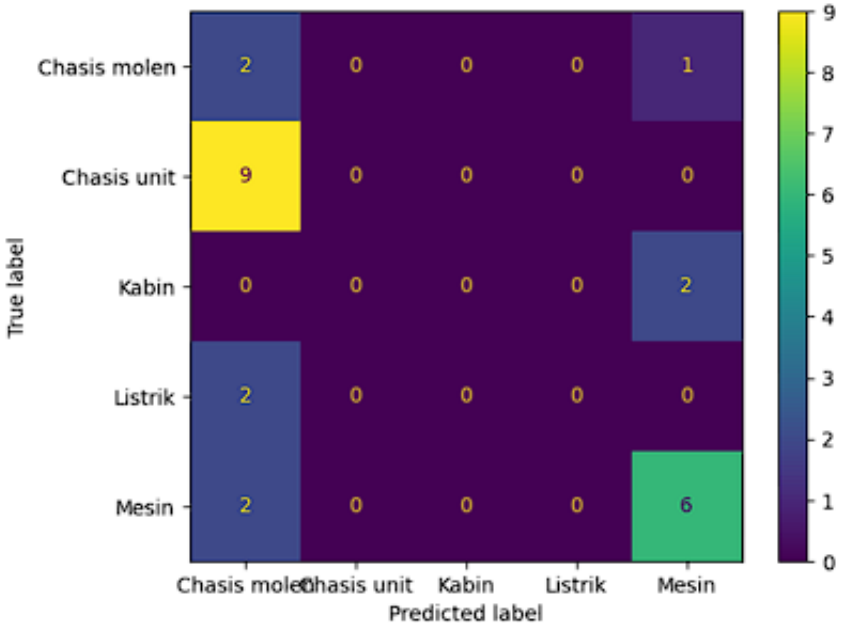


Figure 14. *Confusion Matrix Pada Data Latih 60% dan Data Uji 40%*

Berdasarkan pada gambar 6, perhitungan nilai akurasi dari confusion matrix menggunakan persamaan sebagai berikut:

$$\text{Akurasi} = \frac{\text{Jumlah data terprediksi benar}}{\text{Jumlah seluruh data}} \times 100\%$$

$$\text{Akurasi} = \frac{2+6}{24} \times 100\%$$

$$\text{Akurasi} = 0{,}33$$

Figure 15.

Jenis Kerusakan	Tingkat Risiko	Frekuensi Kerusakan	Bagian Kerusakan (Asli)	Bagian Kerusakan (Prediksi)	Hasil Pengujian
1	1	109	Chasis unit	Chasis molen	Salah
6	2	102	Chasis unit	Chasis molen	Salah
37	2	60	Listrik	Chasis molen	Salah
46	2	37	Mesin	Mesin	Benar
14	1	55	Chasis unit	Chasis molen	Salah
55	2	27	Mesin	Mesin	Benar
34	2	44	Chasis molen	Chasis molen	Benar
49	2	87	Mesin	Chasis molen	Salah
13	2	76	Chasis unit	Chasis molen	Salah
58	2	19	Kabin	Mesin	Salah
47	2	32	Mesin	Mesin	Benar
51	2	13	Mesin	Mesin	Benar
32	1	122	Chasis molen	Chasis molen	Benar
4	2	51	Chasis unit	Chasis molen	Salah
53	2	35	Mesin	Mesin	Benar
18	2	80	Chasis unit	Chasis molen	Salah
9	2	91	Chasis unit	Chasis molen	Salah

Indonesian Journal of Innovation Studies
Vol. 25 No. 3 (2024): July
DOI: 10.21070/ijins.v25i3.1156

7	2	65	Chasis unit	Chasis molen	Salah
41	2	30	Mesin	Mesin	Benar
5	2	135	Chasis unit	Chasis molen	Salah
44	2	113	Mesin	Chasis molen	Salah
20	2	34	Chasis molen	Mesin	Salah
35	2	63	Listrik	Chasis molen	Salah
59	2	8	Kabin	Mesin	Salah

Table 14. *Confusion Matrix Pada Data Latih 60% dan Data Uji 40%*

Pada tabel 14, dapat disimpulkan pada pembagian data 60% data latih dan 40% data uji, diperoleh kelas terbanyak pada hasil prediksi yaitu chasis molen. Hasil kelas terbanyak ini menjadi hasil kerusakan yang diprediksi akan terjadi di masa mendatang dengan akurasi sebesar 33%, hasil akurasi yang diperoleh dibuktikan dengan perbedaan dari hasil data prediksi yang tidak sesuai dengan data asli.

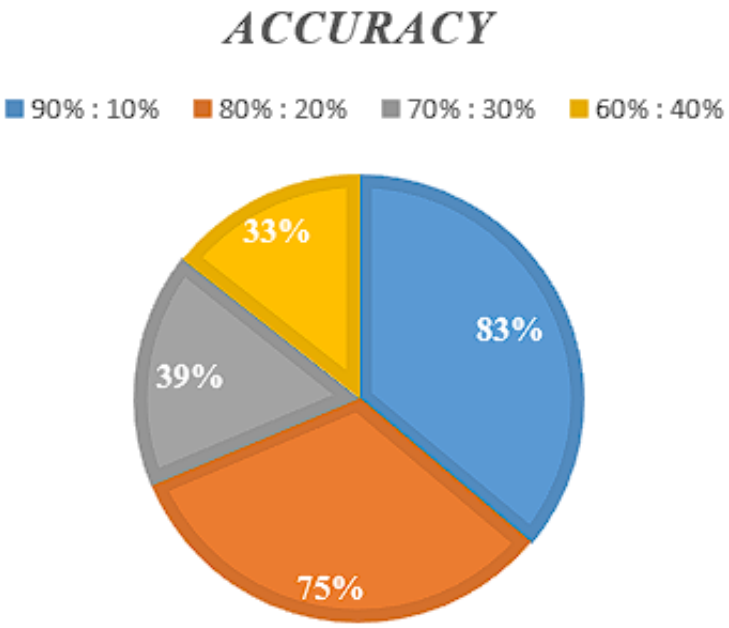


Figure 16. *Hasil Akurasi*

Berdasarkan gambar diatas, performa klasifikasi K-NN yang terbaik menggunakan data latih sebesar 90% dan data uji sebesar 10% dengan nilai k=20, akurasi yang didapat sebesar 83% dengan error sebesar 17%. Berikut adalah rekapitulasi jumlah kerusakan truk mixer berdasarkan bagian kerusakannya periode Januari 2018-Desember 2022, dapat dilihat pada tabel 15.

Bagian Kerusakan	2018	2019	2020	2021	2022
Chasis unit	337	322	365	322	348
Chasis molen	271	273	252	274	265
Listrik	112	104	107	104	110
Mesin	163	184	163	143	159
Kabin	14	13	10	16	12

Table 15. *Rekapitulasi Jumlah Kerusakan Truk Mixer Selama 5 Tahun (Data Asli)*

Berdasarkan tabel diatas, data asli perusahaan menunjukkan bahwa total kerusakan yang banyak terjadi yaitu pada bagian kerusakan chasis unit. Hal ini dapat menjadi acuan dalam membandingkan hasil prediksi yang dilakukan pada setiap variasi dataset. Dalam hal ini data latih 90% dan data uji 10% diprediksi benar dan sesuai dengan tingkat akurasi 83%. Kemudian pada data latih 80% dan data uji 20% diprediksi untuk kerusakan di periode selanjutnya adalah chasis unit dan/ atau mesin dengan akurasi 75%. Kemudian pada data latih 70% dan data uji 30% diprediksi untuk kerusakan di periode selanjutnya adalah chasis molen dengan akurasi 39%. Sedangkan pada data latih 60% dan data uji 40% diprediksi untuk kerusakan di periode selanjutnya adalah chasis molen dengan akurasi 33%.

Simpulan

Sistem yang dirancang dengan menggunakan algoritma K-Nearest Neighbor (K-NN) dapat digunakan dalam deteksi kerusakan truk mixer dengan tingkat akurasi sebesar 83% dan error sebesar 17% dengan nilai k=20 dan data latih sebesar 90% data uji sebesar 10%. Maka, dinyatakan bahwa untuk periode selanjutnya kerusakan yang akan terjadi yaitu pada bagian chasis unit. Hal ini dapat disimpulkan bahwa jumlah data latih yang besar dapat memberikan tingkat akurasi yang tinggi. Dalam penelitian Algoritma K-Nearest Neighbor dengan pendekatan Cross Industry Standard Process for Data Mining (CRISP-DM) dapat menjadi rekomendasi perencanaan perawatan kepada perusahaan maupun usaha lainnya untuk mendeteksi kerusakan truk mixer atau kendaraan lainnya, agar dapat: mengevaluasi interval waktu perawatan kendaraan, untuk meningkatkan availability, merencanakan penyediaan kebutuhan dan sumber daya manusia dalam penjadwalan waktu perawatan agar penanganan kerusakan dapat teratasi dengan cepat dan tepat.

Adapun saran terkait penelitian yang dilakukan, yaitu perlu dilakukannya perbandingan dengan algoritma lainnya, menambah inputan atribut dan dikembangkan dengan menggunakan bahasa pemrograman lain seperti Java, PHP, C++, SQL, dan lain sebagainya.

References

1. [1] C. Dhaniel, M. Pamadi, and A. Savitri, “Project Management of The Icon Housing Development Using Construction Management During the Covid-19 Pandemic,” vol. 5, pp. 548–557, 2022.
2. [2] A. M. Liang and Koespiadi, “Effect of Concrete Material Quality on Construction Cost Efficiency of High-Rise Buildings,” vol. 3, pp. 1–8, 2019.
3. [3] M. H. A. Sarwandy and N. Royan, “Productivity of Backhoe Excavator Equipment on the Al Zafa Housing Project, Tegal Binangun, Palembang City,” pp. 121–125.
4. [4] K. Lorosae, A. I. Sembiring, and S. Debataraja, “Analysis of Heavy Equipment Productivity on Ready Mix Concrete Work: Case Study of Lau Simeme Dam Spillway Building,” vol. 11, no. 1, pp. 95–111, 2023.
5. [5] A. A. W. P. R, F. Rozi, and F. Sukmana, “Unilever Product Sales Prediction Using K-Nearest Neighbor Method,” JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika), vol. 6, no. 1, pp. 155–160, 2021, doi: 10.29100/jipi.v6i1.1910.
6. [6] T. Jaelani, M. Yamin, and C. P. Mahandari, “Machine Learning for Predicting National Sugar Production,” JMPM (Jurnal Material dan Proses Manufaktur), vol. 6, no. 1, pp. 31–36, 2022, doi: 10.18196/jmpm.v6i1.14897.
7. [7] F. M. Subqi and D. Anggraini, “Data Mining for Predictive Maintenance of Production Machines Based on Machine Breakdown Database Using Naïve Bayes Classifier,” Jurnal Ilmiah Komputasi, vol. 20, no. 2, pp. 143–154, 2021, doi: 10.32409/jikstik.20.2.368.
8. [8] M. N. Maskuri, Harliana, K. Sukerti, and R. M. H. Bhakti, “Application of K-Nearest Neighbor (KNN) Algorithm for Predicting Stroke Disease,” Jurnal Ilmiah Intech Information Technology Journal UMUS, vol. 4, no. 1, pp. 130–140, 2022.
9. [9] M. A. Hasanah, S. Soim, and A. S. Handayani, “Implementation of CRISP-DM Model Using Decision Tree Method with CART Algorithm for Predicting Rainfall Potential Flood,”

Jurnal Aplikasi Informatika dan Komputer, vol. 5, no. 2, pp. 103–108, 2021, doi: 10.30871/jaic.v5i2.3200.
10. [10] S. Saepudin, M. Muslih, and Sihabudin, “Major Selection Using K-Nearest Neighbor Method for New Students,” Jurnal Rekayasa Teknologi Nusa Putra, vol. 5, no. 2, pp. 15–19, 2019.
11. [11] I. G. Gusti, M. Nasrun, and R. A. Nugrahaeni, “Recommendation System for Car Selection Using K-Nearest Neighbor (KNN) Collaborative Filtering,” TEKTRIKA - Jurnal Penelitian dan Pengembangan Telekomunikasi, Kendali, Komputer, Elektronika dan Elektrik, vol. 4, no. 1, p. 26, 2019, doi: 10.25124/tektrika.v4i1.1846.
12. [12] A. N. Utomo and N. Lestari, “Detection of Road Damage Using K-NN (K-Nearest Neighbor) Algorithm,” vol. 10, no. 1, pp. 59–66, 2021.
13. [13] Cumel, D. Zamri, Rahmaddeni, and Syamsurizal, “Comparison of Data Mining Methods for Flood Prediction Using Naive Bayes and KNN Algorithms,” SENTIMAS Seminar Nasional Penelitian dan ..., pp. 40–48, 2022. [Online]. Available: https://journal.irpi.or.id/index.php/sentimas/article/view/353 https://journal.irpi.or.id/index.php/sentimas/article/download/353/132.
14. [14] M. Y. R. Rangkuti, M. V. Alfansyuri, and W. Gunawan, “Application of K-Nearest Neighbor (KNN) Algorithm in Predicting and Calculating Accuracy Levels of Weather Data in Indonesia,” Hexagon, vol. 2, no. 2, pp. 11–16, 2021, doi: 10.36761/hexagon.v2i2.1082.
15. [15] B. Y. A. Pratama and H. A. Yuniarto, “Design of Machine Learning Implementation Process in Maintenance Management to Prevent Derating,” J@ti Undip Jurnal Teknik Industri, vol. 16, no. 2, pp. 134–142, 2021, doi: 10.14710/jati.16.2.134-142.
16. [16] A. D. A. Putra and S. Juanita, “Sentiment Analysis on User Reviews of Bibit and Bareksa Applications Using KNN Algorithm,” JATISI (Jurnal Teknik Informatika dan Sistem Informasi), vol. 8, no. 2, pp. 636–646, 2021, doi: 10.35957/jatisi.v8i2.962.
17. [17] F. Hasanah, T. Suprapti, N. Rahaningsih, and I. Ali, “Implementation of K-Nearest Neighbor Algorithm in Determining Books Based on Interests,” Jurnal Accounting Information Systems, vol. 5, no. 1, pp. 102–111, 2022, doi: 10.32627/aims.v5i1.467.
18. [18] P. Putra, A. M. H. Pardede, and S. Syahputra, “Analysis of K-Nearest Neighbor (KNN) Method in Classifying Iris Flower Data,” Jurnal Teknik Informatika Kaputama, vol. 6, no. 1, pp. 297–305, 2022.
19. [19] Ariyadi, “Classification of Bee Species Based on Image Data Using Support Vector Machine Method,” Jurnal Inovasi Penelitian, vol. 1, no. 6, pp. 1065–1070, 2020.
20. [20] A. A. D. Halim and S. Anraeni, “Classification Analysis of Pneumonia Disease Image Dataset Using K-Nearest Neighbor (KNN) Method,” Indonesian Journal of Data Science, vol. 2, no. 1, pp. 01–12, 2021, doi: 10.33096/ijodas.v2i1.23.
21. [21] M. Wahyudi, R. Buaton, and H. Sembiring, “Diagnosis of Online Game Addiction Symptoms Using K-Nearest Neighbor Method,” Seminar Nasional Informasi, vol. 6, no. 3, pp. 106–117, 2022.
22. [22] M. F. Rilwanu, H. Taufikurachman, and F. Huwaidi, “Application of K-Nearest Neighbor Algorithm for Detecting Diabetes Based on Web Application,” vol. 3, no. 1, pp. 145–152, 2022.
23. [23] M. A. Wiratama and W. M. Pradnya, “Optimization of Data Mining Algorithm Using Backward Elimination for Classification of Diabetes Disease,” vol. 11, pp. 1–12, 2022.
24. [24] M. Fansyuri, “Analysis of K-Nearest Neighbor Classification Algorithm in Determining Accuracy of Customer Satisfaction (Case Study of PT. Trigatra Komunikatama),” Humanika Jurnal Ilmu Sosial, Pendidikan, dan Humaniora, vol. 3, no. 1, pp. 29–33, 2020.