

Penerapan *K-Nearest Neighbor* Dengan Metode *Euclidean Distance* Untuk Klasifikasi Tingkat Ketebalan Cat Di PT XYZ

Hendro Gunawan¹, Ahmad Chusyairi²✉, Muhammad Ikhwani Saputra³
^{1,2,3} PJJ Informatika, Universitas Siber Asia, Jakarta, 12550, Indonesia

ahmadchusyairi@lecturer.unsia.ac.id

Abstract

Painting is one of the important processes in the automotive industry, including in leaf spring manufacturing companies as a component of vehicle parts. Proper classification of painting levels has a great effect on the quality of production results. This study discusses the process of primary painting using three main variables, namely paint thickness, air pressure, and viscosity. The method used is the K-Nearest Neighbor (K-NN) algorithm with distance calculation using the Euclidean Distance method. K-NN was chosen for its ability to handle non-linear data and its simplicity in the classification process. The dataset used consists of painting data classified into two classes, namely 'Good' and 'Not Good'. Based on the results of the cross-validation test, it was obtained that the optimal K parameter value was in the range of K = 1 to K = 2 with an accuracy rate of 97,06%. These results show that the model is able to classify data with a high level of accuracy and has good generalization capabilities.

Keywords: K-Nearest Neighbor (K-NN) Algorithm, Euclidean Distance, Classification, K-Value, Machine Learning.

Abstrak

Pengecatan merupakan salah satu proses penting dalam industri otomotif, termasuk pada perusahaan manufaktur pegas (*leaf spring*) sebagai komponen suku cadang kendaraan. Pengklasifikasian tingkat pengecatan secara tepat sangat berpengaruh terhadap kualitas hasil produksi. Penelitian ini membahas proses pengecatan dasar (*primary painting*) dengan menggunakan tiga variabel utama, yaitu ketebalan cat (*thickness*), tekanan udara (*pressure*), dan viskositas cat (*viscosity*). Metode yang digunakan adalah algoritma *K-Nearest Neighbor* (K-NN) dengan perhitungan jarak menggunakan metode *Euclidean Distance*. K-NN dipilih karena kemampuannya dalam menangani data non-linear dan kesederhanaannya dalam proses klasifikasi. Dataset yang digunakan terdiri dari data pengecatan yang diklasifikasikan dalam dua kelas, yaitu 'Good' dan 'Not Good'. Berdasarkan hasil uji validasi silang (*cross-validation*), diperoleh bahwa nilai parameter K optimal berada pada rentang K = 1 hingga K = 2 dengan tingkat akurasi sebesar 97,06%. Hasil ini menunjukkan bahwa model mampu mengklasifikasikan data dengan tingkat akurasi tinggi serta memiliki kemampuan generalisasi yang baik.

Kata Kunci: Algoritma *K-Nearest Neighbor* (K-NN), *Euclidean Distance*, Klasifikasi, Nilai K, Pembelajaran Mesin (*Machine Learning*).

Jutekom is licensed under a Creative Commons Attribution-Share Alike 4.0 International License.



1. Pendahuluan

Klasifikasi merupakan proses mengkategorikan data ke dalam kelas-kelas dengan properti yang sebanding, menggunakan data pelatihan sebagai referensi dan data uji untuk pengujian. Tujuannya adalah untuk memperkirakan kecenderungan dalam data atau item yang tidak terklasifikasi dari kelas-kelas yang belum memiliki label [1]. Dalam klasifikasi, terdapat dua pendekatan utama, yaitu *supervised learning* dan *unsupervised learning*. *Supervised learning* bekerja dengan membentuk fungsi yang menghubungkan *input* dengan *output* yang diharapkan. Beberapa metode yang umum digunakan dalam klasifikasi *supervised learning* meliputi Regresi Logistik, *K-Nearest Neighbor*, *Support Vector Machine*, *Naïve Bayes*, *Decision Tree*, dan *Random Forest* [2]. Algoritma *K-Nearest Neighbor* merupakan metode dasar yang digunakan untuk mengelompokkan objek berdasarkan kedekatannya dengan data pelatihan [3]. Metode *K-Nearest Neighbor* (K-NN) menjadi salah satu metode yang terkenal, tergabung dalam 10 algoritma top *Data Mining* [4]. Data pelatihan dipetakan ke dalam ruang dengan banyak dimensi, di mana setiap dimensi merepresentasikan karakteristik tertentu dari data tersebut. Pemodelan ansambel dan berbasis tetangga adalah dua metodologi pemodelan penting [5].

Penelitian dengan judul Algoritma *K-Nearest Neighbor* (K-NN) dengan *Euclidean* dan *Manhattan* dalam Klasifikasi Kelulusan Siswa yang dilakukan oleh Hidayati dan Hermawan (2021). Hasil diperoleh setelah pengujian 380 data pelatihan dan 163 data pengujian. Sistem akurasi terbaik dicapai pada $K = 7$ dengan nilai 85,28%. Penelitian selanjutnya dengan judul Penerapan Algoritma *K-Nearest Neighbor* dengan *Euclidean Distance* untuk Menentukan Kelompok Uang Kuliah Tunggal Mahasiswa oleh Purwani dkk. (2022). Dimana penentuan kelompok uang tunggal mahasiswa pada 1.650 data verifikasi UKT mahasiswa tahun 2019-2021 diolah dengan data mining menggunakan bantuan *software RStudio*. Hasil penelitian menunjukkan bahwa klasifikasi dengan K-NN dapat diterapkan dalam penentuan UKT mahasiswa. Dengan data *testing* sebanyak 320 mahasiswa terdapat 23 mahasiswa ditentukan mendapat UKT I, 149 UKT II, 129 UKT III, 32 UKT IV, dan 2 mahasiswa mendapat UKT V. Adapun akurasi dari algoritma tersebut yaitu sebesar 87,58% dalam kategori *Good Classification* dengan

nilai K optimal pada K-NN yang didapatkan dengan nilai *K-Fold Cross Validation* adalah $K = 1$. Penelitian selanjutnya dengan judul Klasifikasi Penyakit pada Tanaman Kopi Arabika Menggunakan Metode *K-Nearest Neighbor* (K-NN) Berbasis Citra yang dilakukan sebelumnya oleh Sandri, dkk. (2023). Pengujian K-NN digunakan untuk mengklasifikasikan penyakit tanaman kopi arabika dengan nilai K yang terbaik adalah $K = 11$ menggunakan jarak *euclidean distance* dengan tingkat akurasi dalam penelitian ini sebesar 94%. Berdasarkan kelebihan dan kekurangan algoritma K-NN serta penelitian-penelitian yang telah dilakukan sebelumnya, telah banyak permasalahan yang diselesaikan dengan mengimplementasikan algoritma klasifikasi K-NN. Salah satu masalah yang dapat diselesaikan dengan algoritma ini adalah klasifikasi tingkat ketebalan cat.

Pesatnya pertumbuhan sektor industri berdampak pada terciptanya persaingan antar badan usaha, baik swasta maupun milik negara, yang bergerak di bidang manufaktur dan jasa. Alhasil, berbagai organisasi akan berlomba-lomba memberikan pelayanan terbaik kepada para kliennya. Salah satunya adalah dengan berfokus pada kualitas barang yang dihasilkan [6].

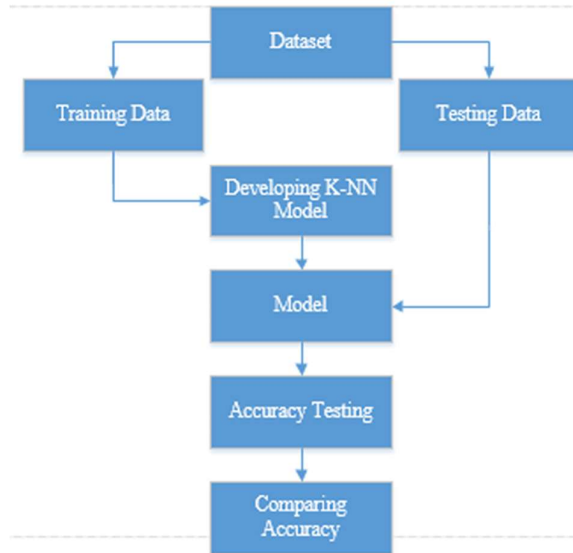
Perkembangan industri otomotif dalam beberapa tahun terakhir telah melahirkan berbagai metode campuran yang lebih efisien dan inovatif, memungkinkan terciptanya komposisi optimal untuk fungsi perlindungan [7]. Namun, berdasarkan observasi yang telah dilakukan didapatkan fakta bahwa masih adanya cacat (*defect*) pada proses cat dasar diantaranya cat yang tidak rata, cat terlalu tipis, cat terlalu tebal, dan cat terkelupas.

2. Metodologi Penelitian

Metode yang digunakan adalah *Euclidean Distance*. Rumus *Euclidean Distance* digunakan untuk menghitung jarak antara data pengujian dan data pelatihan menggunakan algoritma K-NN [8]. Metrik jarak yang paling umum digunakan adalah jarak *Euclidean* [9]. Evaluasi model meliputi akurasi, presisi, *recall*, *f1-score*, dan *accuracy score*. Selain itu, evaluasi juga dilakukan dengan menggunakan *confusion matrix*, dan validasi silang (*cross-validation*).

2.1. Tahapan Penelitian

Berikut ini adalah tahapan penelitian langkah-langkah utama dalam pengembangan dan pengujian model menggunakan algoritma *K-Nearest Neighbor* seperti ditunjukkan pada Gambar 1.



Gambar 1. Tahapan penelitian [10]

2.2. Perhitungan Jarak (*Distance Calculation*)

Jarak *Euclidean* (d) antara dua titik data x (x_1, x_2, \dots, x_n) dan c (c_1, c_2, \dots, c_n) dengan n atribut dijelaskan dalam rumus *Euclidean Distance* sebagai berikut [11]:

$$d = \sqrt{(x_1 - c_1)^2 + (x_2 - c_2)^2 + \dots + (x_n - c_n)^2} \quad (1)$$

Keterangan: dengan d adalah jarak antara dua titik data x dan c , x_n adalah nilai dari data uji (data yang akan diklasifikasikan), dan c_n adalah nilai dari data latih (data pembandingan), dan n adalah jumlah atribut atau variabel yang digunakan.

2.3. Dataset

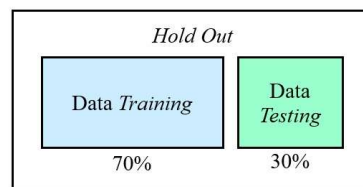
Dataset pada penelitian ini diperoleh dari sebuah instansi industri manufaktur, yaitu PT XYZ yang bergerak di bidang produksi komponen otomotif. Tahapan awal penelitian dilakukan dengan proses pengumpulan data, kemudian dilanjutkan dengan *preprocessing* data yang mencakup integrasi data, pelabelan, transformasi, serta pembersihan data (*data cleaning*) [12]. Seluruh proses pengolahan dan analisis data dilakukan menggunakan *platform Google Colab*. Dataset yang digunakan disimpan dalam file bernama *Dataset.xlsx* dengan total sebanyak 115 entri data. Atribut atau kolom yang terdapat dalam dataset ini meliputi: Nomor, Tanggal, *Type*, Jenis Cat, *Pressure*, *Thickness*,

Viscositas, dan Label. Penjelasan lebih rinci terhadap struktur data adalah sebagai berikut:

- a. Dataset terdiri dari dua label atau kelas kategori, yaitu: *Good* dan *Not Good*.
- b. Atribut *Pressure* (tekanan udara), *Thickness* (ketebalan cat), dan *Viscositas* (kekentalan cat) merupakan variabel independen yang berperan dalam proses klasifikasi dan digunakan dalam perhitungan jarak antar data [13].
- c. Atribut Label adalah variabel dependen atau target *output* yang nilainya dipengaruhi oleh kombinasi dari variabel independen. Variabel ini menunjukkan hasil klasifikasi akhir dari tingkat ketebalan cat, apakah termasuk *Good* atau *Not Good*.

2.4. *Training Data* dan *Testing Data*

Pada tahap ini, dataset dibagi menjadi dua bagian, yaitu data *training* dan data *testing* [14]. Pembagian dilakukan dengan proporsi 70% untuk data *training* dan 30% untuk data *testing*. Dari total 115 data yang tersedia, sebanyak 81 data digunakan untuk proses pelatihan (*training*), sementara 34 data sisanya digunakan untuk proses pengujian (*testing*). Tujuan dari pembagian ini adalah untuk melatih model menggunakan data yang telah diketahui kelasnya, kemudian menguji kemampuan model dalam mengklasifikasikan data baru yang belum dikenali. Proses ini penting untuk mengetahui kemampuan generalisasi model terhadap data yang belum pernah dilihat sebelumnya. Jika hasil pengujian menunjukkan performa model yang kurang optimal, maka penyesuaian terhadap parameter seperti nilai K dapat dilakukan untuk meningkatkan akurasi dan menghindari kesalahan klasifikasi [15].



Gambar 2. *Hold out data* [16]

Mengapa peneliti menggunakan rasio 70%:30% yaitu karena hal-hal sebagai berikut:

1. Keseimbangan antara pelatihan dan pengujian
 - a. Dengan 70% data *training*, model memperoleh informasi yang cukup untuk mempelajari pola dari data.

- b. 30% data testing cukup representatif untuk mengevaluasi generalisasi model terhadap data baru.
- 2. Menghindari *overfitting* atau *underfitting*
 - a. Jika terlalu banyak data digunakan untuk *training* (misalnya 90:10), model bisa terlalu menghafal data dan gagal menggeneralisasinya.
 - b. Jika terlalu sedikit untuk *training* (misalnya 50:50), model tidak memiliki cukup data untuk belajar dengan baik.
- 3. Rasio yang telah teruji secara empiris
 - a. Banyaknya penelitian dan studi terdahulu menggunakan rasio 70:30 dan menunjukkan hasil evaluasi yang stabil dan dapat diandalkan.
 - b. Rasio ini juga direkomendasikan dalam banyak literatur sebagai baseline sebelum mencoba teknik validasi lain seperti *k-fold cross-validation*.
- 4. Sederhana dan efektif

Rasio ini mudah diterapkan, tidak terlalu membebani komputasi, dan cocok untuk dataset berukuran kecil hingga menengah.

2.5. Developing K-NN Model

Proses klasifikasi menggunakan algoritma *K-Nearest Neighbor* (K-NN) dilakukan melalui tiga langkah utama, yaitu: pembangunan model, penerapan model, dan evaluasi model [17]. Pada tahap awal, model dibangun menggunakan data pelatihan (*training data*) yang sudah memiliki atribut dan label kelas yang diketahui. Data ini digunakan untuk "melatih" model agar dapat mengenali pola dan hubungan antar variabel. Setelah model terbentuk, langkah selanjutnya adalah menerapkan model pada data baru (*testing data*) yang belum diketahui kelasnya. Model akan menentukan kelas dari data uji tersebut berdasarkan kemiripan jarak dengan data latih menggunakan metode *Euclidean Distance*. Tahap terakhir adalah evaluasi model, di mana performa model diukur berdasarkan tingkat akurasi klasifikasi terhadap data uji. Evaluasi ini bertujuan untuk mengetahui seberapa baik model mampu menggeneralisasi data baru serta memastikan bahwa model tidak mengalami *overfitting* atau *underfitting*.

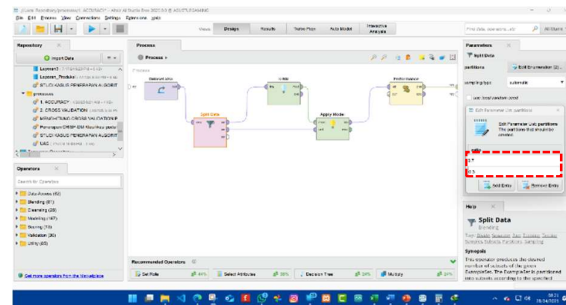
2.6. Model

Model yang digunakan dalam penelitian ini adalah algoritma *K-Nearest Neighbor* (K-NN) dengan pendekatan perhitungan jarak menggunakan metode *Euclidean Distance* [18]. Pemilihan K-NN didasarkan pada kemampuannya dalam

melakukan klasifikasi secara efektif dengan pendekatan yang sederhana namun akurat terhadap data non-linear. Untuk membangun dan menguji model, dataset dibagi menjadi dua bagian, yaitu:

- a. Data Pelatihan (*Training Data*): digunakan untuk membentuk model berdasarkan pola dan hubungan antar atribut.
- b. Data Pengujian (*Testing Data*): digunakan untuk mengevaluasi kinerja model terhadap data yang belum pernah dilihat sebelumnya [19].

Model akan memprediksi kelas suatu data uji dengan membandingkan jaraknya terhadap data pelatihan, kemudian memilih sejumlah K tetangga terdekat untuk menentukan kelas berdasarkan mayoritas suara (*voting*).



Gambar 3. Model

2.7. Accuracy Testing

Performa model K-NN dalam penelitian ini diuji melalui beberapa metrik evaluasi, antara lain: akurasi (*accuracy*), presisi (*precision*), *recall*, *f1-score*, serta *confusion matrix* dan validasi silang (*cross-validation*) [20]. Validasi silang yang digunakan adalah teknik *k-fold cross-validation* dengan nilai K = 10, yang berarti dataset dibagi menjadi 10 bagian yang sama. Model dilatih dan diuji sebanyak 10 kali, masing-masing dengan menggunakan satu bagian sebagai data uji dan sembilan bagian lainnya sebagai data latih. Hasil rata-rata dari 10 pengujian ini digunakan untuk menilai kinerja model secara keseluruhan [21].

Rumus evaluasi performa:

- a. Akurasi (*Accuracy*)

Mengukur rasio prediksi yang benar terhadap keseluruhan data.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

- b. Presisi (*Precision*)

Mengukur ketepatan prediksi positif dari keseluruhan yang diprediksi sebagai positif.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (3)$$

c. *Recall (Sensitivity)*

Mengukur kemampuan model dalam menemukan seluruh kasus positif sebenarnya.

$$\text{Recall} = \frac{TP}{TP + FN} \quad (4)$$

d. *F1-Score*

Rata-rata harmonik dari *precision* dan *recall*, berguna saat data tidak seimbang.

$$\text{F1-Score} = 2 \times \frac{\text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \quad (5)$$

e. *Confusion Matrix*

Matriks yang menunjukkan distribusi prediksi model dibandingkan dengan label sebenarnya. Visualisasi *confusion matrix* ditunjukkan pada Gambar 3.

		True/Actual Class	
		Positive(P)	Negative(N)
Predicted Class	True (T)	True Positive (TP)	False Positive (FP)
	False (F)	False Negative (FN)	True Negative (TN)
		$P=TP+FN$	$N=FP+TN$

Gambar 4. *Confusion Matrix* [22]

2.8. *Comparing Accuracy*

Penelitian ini melakukan perbandingan akurasi dengan menggunakan variasi nilai parameter K pada algoritma *K-Nearest Neighbor* (K-NN). Nilai K divariasikan mulai dari K = 1 hingga K = 18 untuk mengamati pengaruhnya terhadap performa klasifikasi. Penghitungan jarak antar data menggunakan metode *Euclidean Distance*, yaitu metode jarak yang umum digunakan dalam algoritma K-NN karena kesederhanaannya dan kemampuannya menangani data numerik [23]. Data dibagi menggunakan teknik data *splitting* dengan proporsi data latih sebesar 70% dan data uji sebesar 30%, atau dengan parameter *test_size* = 0,3 dan *random_state* = 42.

2.9. *Alat dan Bahan*

Alat dan bahan yang digunakan dalam penelitian ini diantaranya adalah:

1. *Paint Coating Thickness*

Pengukuran ketebalan cat dilakukan dengan menggunakan *coating thickness gauge* [24], yaitu alat pengukur ketebalan yang digunakan untuk memeriksa ketebalan lapisan cat atau pelapis lainnya pada permukaan suatu material.



Gambar 5. *Paint Coating Thickness* [25]

2. *Air Compressor Pressure Gauge*

Pressure gauge merupakan perangkat yang berfungsi untuk mengukur tekanan udara yang dihasilkan oleh kompresor serta tekanan di area mesin produksi.



Gambar 6. *Air Compressor Pressure Gauge* [26]

3. *Iwata Cup*

Iwata cup digunakan sebagai alat pengukur kekentalan cairan (viskositas) pada cat. Menentukan waktu alir (*flow time*) cairan, yaitu waktu yang dibutuhkan cairan untuk keluar dari lubang kecil di bagian bawah. Semakin lama waktu alirannya, semakin tinggi kekentalan cairan tersebut.



Gambar 7. *Iwata Cup*

4. *Stopwatch*

Stopwatch berfungsi untuk mengukur durasi yang dibutuhkan cat agar dapat mengalir melalui *Iwata cup*. Alat ini juga membantu mencatat waktu yang diperlukan cat untuk keluar sepenuhnya dari lubang *viscometer* hingga tetesan terakhir.



Gambar 8. *Stopwatch*

5. *Spray Gun*

Spray gun digunakan untuk penyemprotan yang merata pada proses cat dasar, sehingga lapisan cat dapat menutupi permukaan tanpa celah. Jarak penyemprotan berpengaruh terhadap kualitas akhir pengecatan, sehingga perlu disesuaikan dengan tepat agar hasil pengecatan menjadi optimal [27].



Gambar 9. *Spray Gun* (1. *Paint supply*, 2. *Compressed gas supply*)

3. Implementasi Metode Usulan

3.1. Teknik Pengumpulan Data

Teknik pengumpulan data merupakan langkah penting dalam proses penelitian untuk memperoleh informasi yang akurat dan relevan. Dalam penelitian ini, penulis menggunakan beberapa metode pengumpulan data sebagai berikut:

1. Wawancara

Teknik wawancara dilakukan secara langsung dengan pihak yang memiliki pemahaman mendalam mengenai proses pengecatan di PT XYZ, yaitu tim *Quality Control* (QC). Wawancara dilaksanakan dengan Bapak Muhammad Khafidlin beserta rekan-rekan, guna memperoleh data terkait prosedur, parameter operasional, serta permasalahan yang sering muncul dalam proses pengecatan.

2. Observasi

Observasi dilakukan secara langsung oleh peneliti di lingkungan kerja PT XYZ, khususnya pada area produksi *plant 2* unit 1 bagian *Assembling*. Kegiatan ini difasilitasi oleh Bapak Agus Purnomo selaku shift leader produksi dan didampingi oleh operator bagian pengecatan dasar. Observasi dilaksanakan pada jam istirahat dan setelah jam kerja pada *shift 1*, *shift 2*, dan *shift 3* untuk memastikan akurasi dan konsistensi data yang diamati.

3. Studi Literatur

Studi literatur dilakukan untuk memperkuat landasan teori dan mendukung hasil wawancara dan observasi. Peneliti mengkaji berbagai referensi terpercaya seperti buku, jurnal ilmiah, dan artikel dari situs *web* yang

membahas topik-topik terkait pengecatan, pengendalian kualitas, serta algoritma *K-Nearest Neighbor* dalam pengolahan data industri.

3.2. Teknik Pencatatan, Penyimpanan dan Pengolahan Data

Proses pencatatan, penyimpanan, dan pengolahan data dalam penelitian ini dilakukan melalui beberapa tahapan sebagai berikut:

1. Pencatatan Data

Data awal dicatat secara manual oleh operator di area proses cat dasar menggunakan formulir atau kertas laporan sementara. Pencatatan ini mencakup parameter-parameter penting seperti tekanan udara, ketebalan cat, dan viskositas.

2. Penyimpanan Data

Setelah dicatat secara manual, data tersebut kemudian disalin dan direkap oleh operator ke dalam buku laporan harian produksi. Selanjutnya, data direkam ke dalam format digital menggunakan aplikasi *Microsoft Excel* di komputer bagian produksi. Penyimpanan digital ini memudahkan dalam hal akses, pengelolaan, dan pengolahan data.

3. Pengolahan Data

Data yang telah disimpan dalam bentuk *Excel* kemudian diekstrak dan diproses lebih lanjut menjadi dataset untuk kebutuhan analisis. Pengolahan data dilakukan menggunakan platform *Google Colab*, dan *Rapid Miner* dengan tahapan meliputi integrasi data, pelabelan, pembersihan data (*cleaning*), serta transformasi data agar siap digunakan dalam pemodelan menggunakan algoritma *K-Nearest Neighbor* (K-NN).

3.3. Waktu dan Lokasi Penelitian

Penelitian ini dilaksanakan dengan rincian waktu dan tempat sebagai berikut:

1. Waktu Penelitian

Penelitian dilakukan selama periode 20 November 2024 hingga 04 Januari 2025, yang mencakup tahap pengumpulan data, observasi lapangan, serta implementasi dan pengujian model algoritma.

2. Lokasi Penelitian

Penelitian dilaksanakan di perusahaan PT XYZ yang berlokasi di Gresik, Jawa Timur, khususnya pada departemen produksi *leaf*

spring bagian proses cat dasar (*primary painting*).

3.4. Kebutuhan Perangkat Keras (*Hardware*) dan Perangkat Lunak (*Software*)

Adapun komputer (PC) yang digunakan dalam penelitian ini memiliki spesifikasi sebagai berikut:

1. Prosesor: 12th Gen Intel® Core™ i9-12900KF 3.20 GHz.
2. RAM: 32,0 GB (31,8 GB dapat digunakan).

Selain kebutuhan perangkat keras, diperlukan juga perangkat lunak yang diperlukan untuk penelitian ini. Kebutuhan jenis perangkat lunak yang digunakan meliputi:

1. Sistem Operasi

Sistem operasi komputer yang digunakan dalam penelitian ini adalah *Windows 11 Pro Education* versi 22H2.

2. Aplikasi perangkat yang dipakai

Dalam penelitian ini, penulis memanfaatkan *Google Colab* sebagai platform dalam penerapan metode *K-Nearest Neighbor*.

3. *Microsoft Office Word*

Software ini digunakan untuk menulis laporan hasil penelitian.

4. *Microsoft Excel*

Software ini digunakan untuk menulis dan menyimpan dataset hasil penelitian dalam format *xlsx*.

3.5. Data Latih

Penelitian ini merupakan penelitian eksperimental yang bersifat kuantitatif dan menggunakan algoritma *K-Nearest Neighbor* (K-NN) untuk menganalisis data primer [28]. Analisis dilakukan terhadap data numerik atau kuantitatif yang berasal dari hasil produksi pada proses cat dasar di perusahaan PT XYZ. Dataset pelatihan (*training data*) yang digunakan dalam penelitian ini memiliki beberapa atribut sebagai berikut:

1. No: Merupakan kolom nomor urut data dalam dataset.
2. Tanggal: Merupakan waktu pengambilan dataset.
3. *Type*: Merupakan nama tipe per daun (*leaf spring*) yang diproduksi di PT XYZ.
4. Jenis Cat: Merupakan merk cat yang digunakan dalam proses pengecatan.

5. *Pressure*: Merupakan tekanan udara yang dihasilkan oleh tabung gas kompresor yang diukur dalam satuan bar/psi.

6. *Thickness*: Merupakan tingkat ketebalan cat yang diukur dalam satuan mikrometer (µm).

7. *Viscositas*: Merupakan tingkat keenceran cat yang diukur dengan satuan detik.

8. Label: Kategori hasil klasifikasi berdasarkan hasil *preprocessing* data, terdiri dari dua kelas: *Good* dan *Not Good*. Label ditentukan secara manual berdasarkan ketentuan sebagai berikut:

a. *Pressure*

Good: antara 4 dan 5 bar.
Not Good: < 4 atau > 5 bar.

b. *Thicness*

1. Cat *Zettar* dan *Epoxy*:
Good: > 20 µm.
Not Good: ≤ 20 µm.
2. Cat *Zinc Rich*:
Good: > 15 µm.
Not Good: ≤ 15 µm.

c. *Viscositas*

1. Cat *Zettar*:
Good: antara 8–10 detik.
Not Good: < 8 atau > 10 detik.
2. Cat *Zinc Rich*:
Good: antara 7,5–8,5 detik.
Not Good: < 7,5 atau > 8,5 detik.
3. Cat *Epoxy*:
Good: antara 8,5–9,5 detik.
Not Good: < 8,5 atau > 9,5 detik.
4. Cat Tipe VLD:
Good: antara 14–15 detik.
Not Good: < 14 atau > 15 detik.

Tabel 1. Data Latih

No	Type	Jenis Cat	Pres sure	Thick ness	Visco sitas	Label
1	MSM 03551 #	Zettar	5	22	8	Good
2	HN 00116 # 08	Zinc Rich	3	14	5	Not Good
3	VLD 00007 # 01	Epoxy	5	23	14	Good
....						
112	HN 00116 # 11	Zinc Rich	4	23	8	Good
113	AQR 00009 # 01	Zinc Rich	4	22	8	Good
114	AQR 00009 # 02	Zinc Rich	4	27	8	Good

Tabel 1 berisi data latih yang digunakan dalam proses klasifikasi dengan metode *K-Nearest Neighbor* (KNN).

1. Bagian pertama dari tabel mencantumkan beberapa sampel data yang telah dikategorikan dengan label "Good" atau "Not Good" berdasarkan hasil pengukuran variabel *input* berupa *pressure* (bar), *thickness* (µm), dan *viscositas* (detik).
2. Bagian lanjutan dari tabel berisi tambahan data latih yang memiliki karakteristik serupa dengan bagian pertama, digunakan untuk meningkatkan akurasi model klasifikasi.

3.6. Data Uji

Selanjutnya pada proses pengecatan dilakukan pengukuran secara acak terhadap sebuah produk yang akan dicari labelnya yaitu *type* VLD 00003 # A dengan variabel hasil pengukurannya yaitu tekanan gas kompresor (*pressure*) 5 bar, ketebalan cat (*thickness*) 36 µm, dan tingkat keenceran cat (*viscositas*) 14 detik. Data uji ditunjukkan seperti pada Tabel 2, terlihat bahwa pada data ke-115 belum memiliki label.

Tabel 2. Data Uji

No	Type	Jenis Cat	Pressure	Thickness	Viscositas	Label
1	MSM 03551 # 01	Zettar	5	22	8	Good
2	HN 00116 # 08	Zinc Rich	3	14	5	Not Good
3	VLD 00007 # 01	Epoxy	5	23	14	Good
....						
113	AQR 00009 # 01	Zinc Rich	4	22	8	Good
114	AQR 00009 # 02	Zinc Rich	4	27	8	Good
115	VLD 00003 # A	Epoxy	5	16	14	?

Tabel 2 berisi data uji yang terdiri dari beberapa sampel produk dengan berbagai variabel hasil pengukuran, seperti *pressure* (bar), *thickness* (µm), dan *viscositas* (detik).

1. Bagian pertama dari tabel menunjukkan beberapa sampel produk yang telah diuji dengan label "Good" atau "Not Good" berdasarkan hasil pengukuran.

2. Bagian lanjutan dari tabel menampilkan data tambahan, termasuk data ke-115 (VLD 00003 # A) yang masih belum memiliki label.

3. Keterangan di bawah tabel menjelaskan bahwa data ke-115 masih dalam proses analisis, sehingga labelnya belum ditentukan.

3.7. Menentukan Nilai K

Dalam penelitian ini, dilakukan proses eksplorasi terhadap berbagai nilai K dalam rentang K = 1 hingga K = 18 untuk mencari nilai K yang paling tepat dan optimal. Proses ini dilakukan melalui pengujian performa model pada setiap nilai K menggunakan teknik validasi silang (*cross-validation*) dan metrik akurasi sebagai tolok ukur. Melalui pengujian ini diharapkan dapat ditemukan nilai K yang paling optimal, yaitu nilai yang mampu memberikan keseimbangan antara akurasi tinggi dan kemampuan generalisasi yang baik terhadap data uji.

3.8. Menghitung Jarak Antara Data Baru dan Masing-Masing Data Lainnya

Dalam perhitungan jarak antara tiga titik, digunakan metode *Euclidean Distance* dengan rumus 3D atau rumus ke-(1), sebagai contoh penghitungan jarak data ke-1 hingga data ke-10 adalah sebagai berikut.

Data 1

$$Distance = \sqrt{(5 - 5)^2 + (22 - 36)^2 + (8 - 14)^2} = \sqrt{(0)^2 + (-14)^2 + (-6)^2} = \sqrt{0 + 196 + 36} = \sqrt{232} = 15,2315$$

Data 2

$$Distance = \sqrt{(3 - 5)^2 + (18 - 36)^2 + (5 - 14)^2} = \sqrt{(-2)^2 + (-18)^2 + (-9)^2} = \sqrt{4 + 324 + 81} = \sqrt{409} = 20,2237$$

Data 3

$$Distance = \sqrt{(5 - 5)^2 + (23 - 36)^2 + (14 - 14)^2} = \sqrt{(0)^2 + (-13)^2 + (0)^2} = \sqrt{0 + 169 + 0} = \sqrt{169} = 13,0000$$

Data 4

$$Distance = \sqrt{(5 - 5)^2 + (27 - 36)^2 + (8 - 14)^2} = \sqrt{(0)^2 + (-9)^2 + (-6)^2} = \sqrt{0 + 81 + 36} = \sqrt{117} = 10,8166$$

Data 5

$$Distance = \sqrt{(2 - 5)^2 + (18 - 36)^2 + (5 - 14)^2} = \sqrt{(-3)^2 + (-18)^2 + (-9)^2} = \sqrt{9 + 324 + 81} = \sqrt{414} = 20,3469$$

Setelah menghitung jarak dari data uji ke seluruh data pelatihan, langkah selanjutnya adalah mengurutkan nilai jarak tersebut dari yang paling kecil hingga paling besar, untuk kemudian

memilih sejumlah tetangga terdekat sebanyak K sesuai nilai parameter yang sedang diuji.

3.9. Menentukan Tiga Data Terdekat Dan Label Prediksi Data Baru

Setelah menghitung jarak antara data baru dengan seluruh data pelatihan menggunakan metode *Euclidean Distance*, langkah selanjutnya adalah menentukan K tetangga terdekat untuk proses klasifikasi. Pada contoh ini, digunakan K = 3. Jarak dari setiap data pelatihan ke data baru diurutkan secara *ascending* (dari yang paling dekat ke paling jauh), kemudian dipilih tiga data terdekat. Hasil peringkat jarak dan label dari ketiga data tersebut disajikan pada Tabel 3.

Tabel 3. Data Baru dengan Nilai K = 3

No	Data	Type	Pressure	Thickness	Viscosity	Jarak	Label
1	115	VLD 00003 # A	5	36	14	0,0000	Good
2	65	VLD 00008 # 01	5	36	14	0,0000	Good
3	59	VLD 00008 # 06	5	34	14	2,0000	Good
...							
113	75	MSM 02806 # 52	6	7	7	29,8496	Not Good
114	104	VL 00006 # 02	5	70	14	34,0000	Good
115	97	HN 00116 # 02	4	70	8	34,5398	Good

Tabel 3 menunjukkan klasifikasi data baru menggunakan algoritma *K-Nearest Neighbor* (KNN) dengan K = 3.

- Data baru yang diuji adalah VLD 00003 # A dengan parameter *thickness* 36 μm , *pressure* 5 bar, dan *viscositas* 14 detik.
- Kolom "Jarak" menunjukkan hasil perhitungan *Euclidean Distance* antara data baru dengan data lainnya dalam dataset.
- Tiga data terdekat (K = 3) adalah:
 - Data ke 115 (VLD 00003 # A) dengan jarak 0,0000.
 - Data ke 65 (VLD 00008 # 01) dengan jarak 0,0000.
 - Data ke 59 (VLD 00008 # 06) dengan jarak 2,0000.
- Semua data dalam K = 3 memiliki label *Good*.

Karena ketiga tetangga terdekat semuanya memiliki label "Good", sehingga berdasarkan prinsip *majority voting* dalam algoritma K-NN

maka label prediksi untuk data baru juga ditentukan sebagai: *Good*.

4. Hasil Dan Analisa

4.1. Hasil Akurasi Berdasarkan Nilai K

Akurasi model *K-Nearest Neighbor* (K-NN) untuk setiap nilai parameter K ditunjukkan pada Gambar 9. Berdasarkan hasil pengujian, diperoleh akurasi tertinggi sebesar 97,06% pada nilai K = 1 hingga K = 2. Hal ini menunjukkan bahwa model mampu melakukan klasifikasi data dengan sangat baik pada nilai-nilai tersebut. Namun demikian, akurasi mulai mengalami penurunan pada nilai K > 3, yaitu menjadi 94,12% pada K = 3 hingga K = 15. Penurunan ini mengindikasikan bahwa model mulai melakukan generalisasi dengan baik dan tidak terlalu sensitif terhadap data pelatihan (tidak *overfitting*). Selanjutnya, akurasi kembali menurun menjadi 91,18% pada nilai K = 16 hingga K = 18. Penurunan ini menunjukkan adanya kecenderungan *underfitting*, di mana model terlalu menyamaratakan sehingga kehilangan ketajaman dalam melakukan klasifikasi. Berdasarkan hasil evaluasi tersebut, dapat disimpulkan bahwa K = 1 dan K = 2 memiliki akurasi yang sangat baik dalam mengklasifikasikan data. Berdasarkan hasil evaluasi tersebut, dapat disimpulkan bahwa K = 1 dan K = 2 memiliki akurasi yang sangat baik dalam mengklasifikasikan data. Dengan demikian, meskipun nilai K = 1 memberikan akurasi paling tinggi, pemilihan nilai K dalam rentang 6 hingga 11 lebih direkomendasikan untuk menghasilkan model yang lebih stabil, andal, dan tidak terlalu bergantung pada data pelatihan.

Tabel 4. Akurasi untuk Setiap Nilai K

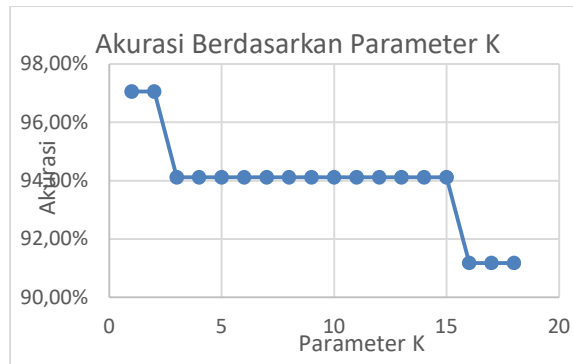
Parameter K	Akurasi
K = 1	97,06%
K = 2	97,06%
K = 3	94,12%
K = 4	94,12%
K = 5	94,12%
K = 6	94,12%
K = 7	94,12%
K = 8	94,12%
K = 9	94,12%
K = 10	94,12%
K = 11	94,12%
K = 12	94,12%
K = 13	94,12%
K = 14	94,12%
K = 15	94,12%
K = 16	91,18%
K = 17	91,18%
K = 18	91,18%

4.2. Grafik Nilai K Terhadap Akurasi

Grafik pada Gambar 10 menunjukkan hubungan antara nilai parameter K dan tingkat akurasi model *K-Nearest Neighbor* (K-NN) yang dihasilkan. Sumbu horizontal (X) merepresentasikan variasi nilai K dari 1 hingga 18, sementara sumbu vertikal (Y) menunjukkan nilai akurasi dari model untuk setiap K. Berdasarkan grafik:

1. Pada nilai K = 1 hingga K = 2, akurasi tetap berada pada 97,06%, menunjukkan model mampu mengklasifikasikan dengan sempurna.
2. Mulai dari K = 3 hingga K = 15, terjadi penurunan akurasi ke 94,12%, namun masih menunjukkan performa yang sangat baik.
3. Pada K ≥ 16, akurasi menurun lebih lanjut menjadi 91,18% dan tetap stabil hingga K = 18, menunjukkan kecenderungan *underfitting*.

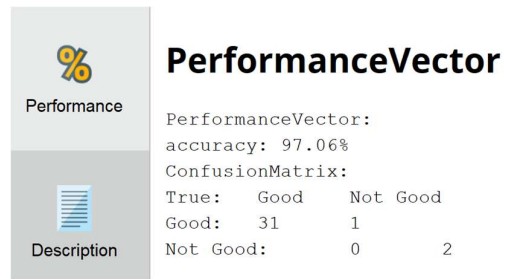
Dari grafik ini dapat disimpulkan bahwa nilai K terbaik adalah dalam rentang 1 hingga 2 karena mampu memberikan akurasi tinggi sekaligus menjaga kemampuan generalisasi model.



Gambar 10. Rentang Nilai K Optimal

4.3. Confusion Matrix

Confusion matrix pada Gambar 11 digunakan untuk mengevaluasi performa model klasifikasi dengan memberikan gambaran menyeluruh terhadap hasil prediksi dibandingkan dengan data sebenarnya. *Confusion matrix* menyusun data uji ke dalam kategori *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN), sehingga memudahkan dalam mengukur akurasi dan kesalahan model.



Gambar 11. Confusion Matrix

Berdasarkan hasil pada Gambar 11, diketahui bahwa:

1. Sebanyak 31 data kelas 'Good' berhasil diklasifikasikan dengan benar sebagai positif (*True Positive*).
2. Sebanyak 2 data kelas 'Not Good' berhasil diklasifikasikan dengan benar sebagai negatif (*True Negative*).
3. Terdapat kesalahan klasifikasi, sebagai berikut:
 - a. *False Positive* (FP): Tidak ada data negatif yang salah diklasifikasikan sebagai positif.
 - b. *False Negative* (FN): Terdapat 1 data positif yang salah diklasifikasikan sebagai negatif.

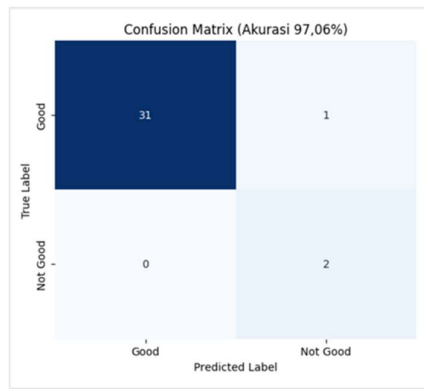
$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} = \frac{31+2}{31+2+0+1} = \frac{33}{34} =$$

$$0,97058 \times 100\% = 97,06\%$$

Hasil ini menunjukkan bahwa model K-NN memiliki performa klasifikasi yang sangat baik pada nilai K terbaik, dengan tingkat kesalahan klasifikasi yang sangat minimal.

4.4. Visualisasi Confusion Matrix

Dengan nilai *False Positive* (FP) = 0 dan *False Negative* (FN) = 1, model menunjukkan klasifikasi yang sangat baik, meskipun terdapat satu kesalahan dalam mengklasifikasikan kelas positif. Secara keseluruhan, akurasi model mencapai 97,06%, menandakan bahwa model mampu mengklasifikasikan data dengan tingkat keakuratan yang tinggi terhadap dataset yang digunakan. Hasil visualisasi *confusion matrix* ditampilkan pada Gambar 12 berikut.



Gambar 12. Visualisasi *Confusion Matrix*

4.5. Kinerja Model

Kinerja model ditunjukkan pada Gambar 13. Untuk kelas *Good*, model memiliki nilai *precision* sebesar 0,97 yang berarti bahwa 97% prediksi untuk kelas *Good* adalah benar. Untuk kelas *Not Good*, nilai *precision* sebesar 0,67 menunjukkan bahwa hanya 67% prediksi untuk kelas tersebut yang benar. Pada metrik *recall*, kelas *Good* memperoleh nilai 0,97 yang berarti hampir semua data yang termasuk dalam kelas *Good* berhasil diklasifikasikan dengan benar. Untuk kelas *Not Good*, nilai *recall* sebesar 0,67 menunjukkan bahwa hanya 67% data yang sebenarnya *Not Good* berhasil dikenali dengan benar. Nilai *f1-score* untuk kelas *Good* adalah 0,97 sedangkan untuk kelas *Not Good* adalah 0,67 mencerminkan keseimbangan antara *precision* dan *recall* pada kedua kelas tersebut. Rata-rata keseluruhan kinerja model adalah:

1. *Macro average: precision* 0,82, *recall* 0,82, dan *f1-score* 0,82.
2. *Weighted average: precision* 0,82, *recall* 0,82, dan *f1-score* 0,82.

Akurasi keseluruhan model sebesar 0,94 (94%) menunjukkan bahwa model mampu mengklasifikasikan data dengan tingkat ketepatan yang tinggi, meskipun terdapat ketidakakuratan pada prediksi kelas *Not Good*.

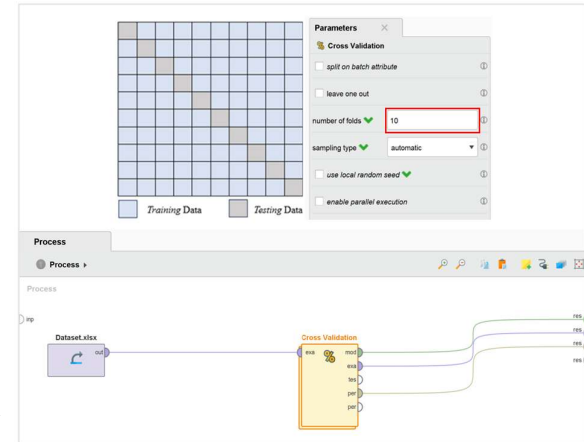
	precision	recall	f1-score	support
Good	0.97	0.97	0.97	32
Not Good	0.67	0.67	0.67	3
accuracy			0.94	35
macro avg	0.82	0.82	0.82	35
weighted avg	0.94	0.94	0.94	35

Gambar 13. *Classification Report*

4.6. Cross Validation

Untuk mengevaluasi stabilitas dan generalisasi model dalam melakukan klasifikasi, digunakan

metode *k-fold cross-validation* dengan jumlah lipatan (*K*) sebanyak 10 seperti ditunjukkan pada Gambar 14. Proses ini dilakukan dengan membagi data menjadi 10 bagian, di mana pada setiap iterasi satu bagian digunakan sebagai data uji, sedangkan sembilan bagian lainnya digunakan sebagai data latih.



Gambar 14. *Cross Validation K-Fold = 10*

4.7. Tingkat Ketebalan Cat

Dalam analisis ini, dataset yang digunakan berisi informasi mengenai berbagai tipe cat dasar beserta nilai ketebalan (*thickness*) dalam satuan mikrometer (μm). Untuk menentukan cat dengan ketebalan tertinggi dan terendah, digunakan fungsi *idxmax()* dan *idxmin()* dalam *python*, yang masing-masing mencari indeks dengan nilai maksimum dan minimum dari kolom *thickness*. Hasil dari analisis ini menunjukkan bahwa tipe HN 00116 # 02 memiliki ketebalan tertinggi sebesar 70 μm , sedangkan tipe MSM 03521 # 03 memiliki ketebalan terendah sebesar 7 μm seperti ditunjukkan pada Gambar 17.

```
[ ] 1 # Baca dataset
2 dataset = pd.read_excel("Dataset.xlsx")
3
4 # Cari cat paling tebal
5 most_thick = dataset.loc[dataset["Thickness"].idxmax()]
6 most_thin = dataset.loc[dataset["Thickness"].idxmin()]
7
8 # Hasil
9 print("1. Cat paling tebal:")
10 print(f"Type: {most_thick['Type']}, Thickness: {most_thick['Thickness']} \u00b5m")
11
12 print("\n2. Cat paling tipis:")
13 print(f"Type: {most_thin['Type']}, Thickness: {most_thin['Thickness']} \u00b5m")
```

1. Cat paling tebal:
Type: HN 00116 # 02, Thickness: 70 μm

2. Cat paling tipis:
Type: MSM 03521 # 03, Thickness: 7 μm

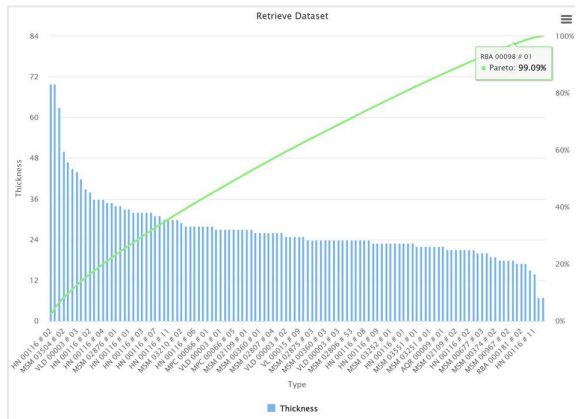
Gambar 15. Kode dan Hasil Analisis Ketebalan Cat Menggunakan *Python*

4.8. Diagram Pareto

Gambar 18 di bawah ini merupakan *Pareto Chart* yang menunjukkan ketebalan cat (*thickness*) berdasarkan tipe (*type*). Berikut adalah penjelasannya:

1. Batang berwarna biru (*histogram*): menunjukkan nilai ketebalan cat untuk setiap tipe. Data telah diurutkan dari nilai ketebalan tertinggi hingga terendah, sehingga memudahkan identifikasi tipe-tipe cat yang paling dominan berdasarkan ketebalannya.
2. Garis berwarna hijau (*kurva pareto*): merepresentasikan persentase kumulatif ketebalan dari seluruh tipe cat. Persentase ini ditampilkan pada sumbu vertikal kanan (0%-100%). Kurva ini membantu dalam menentukan kontribusi proporsional dari setiap tipe terhadap total ketebalan cat.

Sebagai contoh, pada grafik ditunjukkan bahwa tipe RBA 00098 # 01 berkontribusi pada titik kumulatif sekitar 99,09%, yang menunjukkan bahwa sebagian besar ketebalan cat didominasi oleh beberapa tipe cat saja. Penggunaan *Pareto Chart* ini sangat bermanfaat untuk analisis prioritas dalam pengendalian kualitas maupun optimalisasi produksi, khususnya dalam menentukan tipe cat yang paling signifikan terhadap parameter ketebalan.



Gambar 16. Diagram Pareto Ketebalan Cat

5. Kesimpulan

5.1. Kesimpulan

Berdasarkan proses perancangan, implementasi, serta pengujian yang telah dilakukan dalam penelitian terkait ketebalan cat di PT XYZ, dapat disimpulkan bahwa:

1. Algoritma *K-Nearest Neighbor* (K-NN) berhasil diterapkan untuk mengklasifikasikan tingkat

ketebalan cat pada proses *primary painting*. Dengan menggunakan metode *Euclidean Distance*, model mampu menghitung jarak antar sampel data dan menentukan kelas berdasarkan sejumlah tetangga terdekat (*nearest neighbor*).

2. Efektivitas K-NN dalam klasifikasi dengan rasio data latih 70% dan data uji 30%, model menunjukkan performa tinggi dengan tingkat kesalahan yang relatif kecil. Hal ini membuktikan bahwa K-NN cukup efektif dalam mengklasifikasikan tingkat ketebalan cat.
3. Berdasarkan hasil uji validasi silang (*cross-validation*), diperoleh bahwa nilai parameter K optimal berada pada rentang K = 1 hingga K = 2 dengan tingkat akurasi sebesar 97,06%. Hasil ini menunjukkan bahwa model mampu mengklasifikasikan data dengan tingkat akurasi tinggi serta memiliki kemampuan generalisasi yang baik.
4. Hasil analisis menunjukkan bahwa variabel tekanan udara (*pressure*), ketebalan cat (*thickness*), dan viskositas cat (*viscosity*) memiliki pengaruh yang signifikan terhadap performa klasifikasi. Ketiga variabel ini berperan penting dalam menentukan hasil klasifikasi ketebalan cat oleh model K-NN, karena nilai-nilai yang terkandung di dalamnya menjadi dasar dalam penghitungan jarak antar data.

5.2. Saran

Berikut adalah beberapa saran untuk penelitian ini agar dapat dikembangkan lebih lanjut:

1. Saat ini, pemilihan nilai K optimal dilakukan dengan *cross-validation*. Untuk meningkatkan akurasi, dapat dicoba *Grid Search* atau *Random Search* agar pemilihan nilai K lebih optimal dan tidak *overfitting* pada dataset tertentu.
2. Selain itu, eksplorasi metode lain seperti *Weighted K-NN*, di mana kontribusi tetangga lebih dekat diberi bobot lebih tinggi, bisa menjadi alternatif.
3. Selain *Euclidean Distance*, bisa diuji metode lain seperti *Manhattan Distance*, *Minkowski Distance*, atau *Mahalanobis Distance* untuk melihat apakah ada peningkatan akurasi dalam klasifikasi ketebalan cat.
4. Hasil klasifikasi menggunakan K-NN dapat dibandingkan dengan algoritma lain seperti *Decision Tree*, *Random Forest*, atau *Support Vector Machine* (SVM) untuk melihat apakah

- ada metode yang lebih baik dalam klasifikasi ketebalan cat.
5. Meningkatkan jumlah sampel dan keragaman dataset (misalnya variasi lebih banyak dalam tipe cat, tekanan udara, atau suhu lingkungan) agar model lebih generalisasi dan *robust* terhadap kondisi dunia nyata.
 6. Selain *pressure*, *viscosity*, dan *thickness*, faktor eksternal lain seperti suhu lingkungan, kelembapan udara, atau jenis material dasar dapat diteliti untuk melihat apakah memiliki dampak terhadap kualitas pengecatan.
 7. Menggunakan metode *Feature Selection* seperti *Principal Component Analysis* (PCA) atau *Recursive Feature Elimination* (RFE) untuk menentukan fitur yang paling berpengaruh dalam prediksi ketebalan cat, sehingga model bisa lebih efisien.
 8. Model yang dikembangkan dapat diintegrasikan ke dalam sistem berbasis IoT atau *dashboard monitoring real-time* yang dapat membantu operator dalam mengontrol ketebalan cat selama proses produksi.

Daftar Rujukan

- [1] [1] R. Putri Fitrianti, A. Kurniawati, D. Agusten, J. Sistem Informasi, and F. Ilmu Komputer dan Teknologi Informasi, "Implementasi Algoritma K-Nearest Neighbor Terhadap Analisis Sentimen Review Restoran Dengan Teks Bahasa Indonesia," *Seminar Nasional Aplikasi Teknologi Informasi (SNATI)*, pp. 1907-5022, 2019, [Online]. Available: <https://www.google.com>,
- [2] [2] F. S. Pamungkas, B. D. Prasetya, and I. Kharisudin, "Perbandingan Metode Klasifikasi Supervised Learning pada Data Bank Customers Menggunakan Python," *PRISMA, Prosiding Seminar Nasional Matematika*, vol. 3, pp. 689-694, 2019, [Online]. Available: <https://journal.unnes.ac.id/sju/index.php/prisma/>
- [3] [3] N. A'yuni Ramadhani and H. A. Rosyid, "Algoritma-Algoritma Data Mining untuk Klasifikasi Data," *Jurnal Inovasi Teknik dan Edukasi Teknologi*, vol. 2, no. 12, pp. 550-556, Dec. 2022, doi: 10.17977/um068v2i122022pxxx-xxx.
- [4] [4] F. Kurnia, J. Kurniawan, I. S. Fahmi, and S. Monalisa, "Klasifikasi Keluarga Miskin Menggunakan Metode K-Nearest Neighbor Berbasis Euclidean Distance," *Seminar Nasional Teknologi Informasi, Komunikasi dan Industri (SNTIKI)*, vol. 12, pp. 2579-5406, Nov. 2019.
- [5] [5] F. Abbas *et al.*, "Landslide susceptibility assessment along the Karakoram highway, Gilgit Baltistan, Pakistan: A comparative study between ensemble and neighbor-based machine learning algorithms," *Science of Remote Sensing*, vol. 9, Jun. 2024, doi: 10.1016/j.srs.2024.100132.
- [6] [6] Z. Rozikin, A. T. Zy, and A. Z. Kamalia, "Prediksi Ketebalan Powder Coating Menggunakan Algoritma SVM Dan Naïve Bayes," *Jurnal Informasi Teknologi*, vol. 4, pp. 226-231, 2023, doi: 10.47065/bit.v3i1.
- [7] [7] F. B. S. T. W. R. S. D. R. Rani Anggrainy, "Sintesis Cat Epoxy Dengan Variasi Suhu Pengeringan," *Jurnal Perancangan, Manufaktur, Material, Dan Energi*, pp. 42-49, May 2023.
- [8] [8] L. Rahmawati and A. Dwi Indriyanti, "Penerapan Metode K-Nearest Neighbor (K-NN) Untuk Prediksi Penjualan Pakaian (Studi Kasus: UMKM Kresna)," *Journal of Emerging Information Systems and Business Intelligence*, pp. 307-313, 2024.
- [9] [9] A. Sumayli, "Development of advanced machine learning models for optimization of methyl ester biofuel production from papaya oil: Gaussian process regression (GPR), multilayer perceptron (MLP), and K-nearest neighbor (KNN) regression models," *Arabian Journal of Chemistry*, vol. 16, no. 7, Jul. 2023, doi: 10.1016/j.arabjc.2023.104833.
- [10] [10] M. Zhu *et al.*, "Prediction of constitutive model for basalt fiber reinforced concrete based on PSO-KNN," *Heliyon*, vol. 10, no. 11, Jun. 2024, doi: 10.1016/j.heliyon.2024.e32240.
- [11] [11] A. Chusyairi, "Clustering Data Cuaca Ekstrim Indonesia dengan K-Means dan Entropi," *Journal of Informatics and Communications Technology*, vol. 5, no. 1, pp. 1-010, Jun. 2023, doi: 10.52661.
- [12] [12] T. Gori, A. Sunyoto, and H. Al Fatta, "Preprocessing Data dan Klasifikasi untuk Prediksi Kinerja Akademik Siswa," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 11, no. 1, pp. 215-224, Feb. 2024, doi: 10.25126/jtiik.20241118074.
- [13] [13] O. Marlina Susianti, "Perumusan Variabel Dan Indikator Dalam Penelitian Kuantitatif Kependidikan," *Jurnal Pendidikan Rokania*, vol. 9, p. 18, 2024, doi: 10.37728/jpr.v9i1.1066.
- [14] [14] S. Aisyah, S. Wahyuningsih, and fdt Amijaya, "Peramalan Jumlah Titik Panas Provinsi Kalimantan Timur Menggunakan Metode Radial Basis Function Neural Network," *Jambura Journal of Probability and Statistics*, vol. 2, no. 2, pp. 64-74, Nov. 2021, doi: 10.34312/jjps.v2i2.10292.
- [15] [15] A. S. Ritonga and I. Muhandhis, "Teknik Data Mining Untuk Mengklasifikasikan Data Ulasan Destinasi Wisata Menggunakan Reduksi Data Principal Component Analysis (PCA)," *Jurnal Ilmiah Edutic*, vol. 7, no. 2, May 2021.
- [16] [16] P. Purwono, A. Wirasto, and K. Nisa, "Comparison of Machine Learning Algorithms for Classification of Drug Groups," *Sisfotenika*, vol. 11, no. 2, p. 196, Jul. 2021, doi: 10.30700/jst.v11i2.1134.
- [17] [17] D. Azzahra Nasution, H. H. Khotimah, and N. Chamidah, "Perbandingan Normalisasi Data Untuk Klasifikasi Wine Menggunakan Algoritma K-NN," *CESS (Journal of Computer Engineering System and Science)*, vol. 4, no. 1, pp. 2502-7131, Jan. 2019.
- [18] [18] N. Hidayati and A. Hermawan, "K-Nearest Neighbor (K-NN) algorithm with Euclidean and Manhattan in classification of student graduation," *Journal of Engineering and Applied Technology*, vol. 2, no. 2, Aug. 2021, doi: 10.21831/jeatech.v2i2.42777.
- [19] [19] J. Mahasiswa and A. Samudra, "Penerapan Data Mining Untuk Memprediksi Harga Bahan Pangan di Indonesia Menggunakan Algoritma K-Nearest Neighbor," *Jurnal Mahasiswa Akuntansi Samudra (JMAS)*, vol. 4, no. 4, pp. 223-235, Aug. 2023.
- [20] [20] S. Dewi Marliyana, "Uji Performa Spektrofotometer Serapan Atom Thermo Ice 3000 Terhadap Logam Pb Menggunakan CRM 500 dan CRM 697 di UPT

- [21] Laboratorium Terpadu UNS, " *Indonesian Journal Of Laboratory*, vol. 4, no. 2, p. 4887, 2021.
- [22] [21] A. Hagi and D. B. Rarasati, "Sentiment Analysis of Sirekap Application Review Using Logistic Regression Algorithm," *Jurnal Informatika*, vol. 11, no. 2, pp. 55-64, Aug. 2024, doi: 10.31294/inf.v11i2.22066.
- [23] [22] Y. Cnn ... | Brianorman and D. Utami, "Comparative Analysis of CNN Architectures for SIBI Image Classification," vol. 12, no. 1, pp. 61-70, 2024, [Online]. Available: <https://bit.ly/3trwIIH>
- [24] [23] A. Setiawan, "Perbandingan Penggunaan Jarak Manhattan, Jarak Euclid, dan Jarak Minkowski dalam Klasifikasi Menggunakan Metode KNN pada Data Iris," *Jurnal Sains dan Edukasi Sains*, vol. 5, no. 1, pp. 28-37, May 2022, doi: 10.24246/juses.v5i1p28-37.
- [25] [24] A. Ahmad Regal Saputra, "Pengaruh Kekuatan Bonding Terhadap Waktu Pengeringan dan Ketebalan Lapisan Cat Pada Plat Baja Karbon," *Journal of Technical Engineering*, vol. 6, no. 2, pp. 35-43, 2023.
- [26] [25] N. Islahudin, "Teknologi Proses Pengecatan Menggunakan Sistem Atomasi Pada Produk Berbahan Plastik Di Industri Perakitan Sepeda Motor," *Jurnal Ilmiah Teknik Mesin*, vol. 13, no. 1, pp. 15-25, 2019, [Online]. Available: <http://jurnal.umj.ac.id/index.php/sintek>
- [27] [26] D. Suhendra, P. Dan, and Y. Kusuma, "Analisis Kinerja Sistim Kompresor Udara di Jalur Produksi PT. X Melalui Audit Energi," *Jurnal Teknik Mesin*, vol. 10, no. 2, 2021, [Online]. Available: www.en.us.fluke.com
- [28] [27] M. Idra, "Pengaruh Jarak Penyemprotan Spray Gun Merk Meiji Dan Kentaro Terhadap Kualitas Hasil Pengecatan Dengan Tipe Cat Solid," *Automotive Science and Education Journal*, 2020, [Online]. Available: <http://journal.unnes.ac.id/sju/index.php/asej>
- [29] [28] Y. Wicaksono, A. Akbar Harahap, and T. Rochmadi, "Implementasi Business Intelligence untuk Visualisasi Data Pengembangan Kompetensi Aparatur Sipil Negara Di Kabupaten Lebak," *Jurnal Teknologi dan Ilmu Komputer (JUTEKOM)*, Apr. 2024, doi: 10.35134/Jutekom.v9i2.1.