

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Batik merupakan salah satu bentuk seni dan budaya yang kaya di Indonesia, di mana setiap motifnya menyimpan makna dan cerita yang mendalam. Motif batik keberagaman mencerminkan identitas budaya dari berbagai daerah di Indonesia, menjadikannya sebagai warisan yang sangat berharga. Namun, seiring dengan perkembangan teknologi dan *digitalisasi*, muncul tantangan baru dalam melestarikan dan mempromosikan batik. Salah satu tantangan tersebut adalah bagaimana cara mengidentifikasi dan menganalisis berbagai motif batik secara efektif. Pentingnya pengolahan citra, yang memungkinkan kita untuk melakukan analisis mendalam terhadap motif batik dengan memanfaatkan teknik-teknik *modern*.

Batik sendiri merupakan salah satu dari warisan budaya Indonesia, yang bahkan *UNESCO (United Nations Educational, Scientific and Cultural Organization)* telah menetapkan batik merupakan Warisan Kemanusiaan untuk Budaya Non Bendawi (*Masterpieces of the Intangible Cultural Heritage of Humanity*) bagi Indonesia pada tanggal 2 Oktober 2009 Tantangan dalam analisis motif batik adalah segmentasi citra, yaitu proses memisahkan bagian-bagian citra yang memiliki kemiripan fitur visual. Dalam konteks batik, segmentasi berbasis tekstur menjadi metode yang tepat karena motif batik sangat bergantung pada pola dan struktur tekstur yang kompleks.[1]Salah satu metode yang menjanjikan untuk segmentasi citra adalah algoritma *K-Means Clustering*. Metode ini berfungsi

dengan cara mengelompokkan data berdasarkan kesamaan fitur, seperti tekstur, warna, dan pola.

Masalah utama yang melatarbelakangi penelitian “Implementasi Algoritma *K-Means Clustering* untuk Segmentasi Citra Motif Batik Berdasarkan Tekstur” adalah kompleksitas motif batik yang memiliki variasi tekstur rumit, dengan pola dan warna yang sering kali tumpang tindih serta tingkat keteraturan yang berbeda-beda. Kondisi ini membuat proses identifikasi, pengelompokan, dan segmentasi motif batik secara manual menjadi sulit, memakan waktu, dan rawan kesalahan. Selain itu, belum banyak penelitian yang fokus pada segmentasi motif batik berbasis tekstur menggunakan metode yang mampu memberikan hasil pengelompokan akurat secara otomatis.

Penelitian ini terletak pada kebutuhan akan teknologi yang dapat memproses citra batik secara efisien untuk kepentingan pelestarian budaya, dokumentasi digital, dan pengembangan desain. Dengan mengombinasikan *GLCM* sebagai metode ekstraksi fitur tekstur dan *K-Means Clustering* untuk proses pengelompokan, penelitian ini diharapkan dapat menghasilkan segmentasi yang akurat dan konsisten. Hasilnya tidak hanya berkontribusi pada pengembangan ilmu di bidang pengolahan citra dan analisis data, tetapi juga memiliki penerapan praktis seperti katalogisasi digital, serta inspirasi pembuatan desain batik baru yang sesuai dengan karakteristik tekstur yang diinginkan.

Dalam penelitian “Implementasi Algoritma *K-Means Clustering* untuk Segmentasi Citra Motif Batik Berdasarkan Tekstur”, metode *GLCM* (*Gray Level Co-occurrence Matrix*) berperan sebagai tahap ekstraksi fitur tekstur dengan

menghitung hubungan keabuan antar piksel pada jarak dan sudut tertentu untuk menghasilkan parameter seperti *contrast*, *correlation*, *energy*, dan *homogeneity* yang merepresentasikan karakteristik motif batik. Hasil ekstraksi ini berupa vektor fitur numerik yang menjadi input bagi algoritma *K-Means Clustering*, yang kemudian berfungsi mengelompokkan citra batik ke dalam cluster berdasarkan kemiripan fitur teksturnya. Proses *K-Means* dimulai dari inisialisasi centroid secara acak, penghitungan jarak *Euclidean* setiap vektor fitur terhadap centroid, penentuan keanggotaan cluster, hingga pembaruan centroid secara iteratif sampai hasil *konvergen*. Dengan kombinasi ini, *GLCM* menyediakan deskripsi tekstur yang akurat, sedangkan *K-Means* melakukan pengelompokan motif secara efektif untuk mendukung segmentasi citra batik berbasis tekstur. *K-Means* bertugas mengelompokkan motif batik yang memiliki karakteristik tekstur mirip, misalnya motif kasar, motif halus, atau motif dengan pola tertentu.

Algoritma *K-Means Clustering*, yang merupakan salah satu metode *unsupervised learning*, dianggap efektif dalam mengelompokkan piksel citra berdasarkan kemiripan fitur tekstur motif batik di Indonesia.[2] Oleh karena itu, algoritma ini sangat relevan untuk diterapkan dalam konteks ini. Berdasarkan permasalahan yang ada, saya tertarik untuk mengangkat judul skripsi “Implementasi Algoritma *K-Means Clustering* untuk Segmentasi Citra Motif Batik Berdasarkan Tekstur.” Melalui penelitian ini, saya berharap dapat memberikan kontribusi terhadap pengembangan teknologi pengolahan citra batik berbasis komputer.

Segmentasi citra batik berdasarkan tekstur menawarkan beberapa manfaat

signifikan. Otomatisasi proses segmentasi, dibandingkan dengan metode manual yang membutuhkan keahlian khusus, meningkatkan efisiensi dan kecepatan. Lebih lanjut, teknologi ini berkontribusi pada pelestarian budaya batik dengan menyediakan metode digital untuk identifikasi dan pengelompokan motif batik .[3]

Segmentasi citra merupakan tahapan yang pertama kali dilakukan sebelum tahapan analisis citra dalam proses pengenalan citra dari suatu *inputan* tertentu. Fungsi dari segmentasi citra yaitu membagi citra kedalam wilayah-wilayahnya (*region*) berdasarkan kesamaan yang dimilikinya baik berupa tekstur, warna, bentuk, dan lain sebagainya. Penerapan segmentasi citra yang biasa dilakukan antara lain dapat dijumpai dalam aplikasi pengenalan wajah, deteksi kualitas buah, industri makanan kaleng, industri batik, dan lain sebagainya.[4]

Salah satu algoritma populer untuk segmentasi adalah *K-Means Clustering*, sebuah metode *unsupervised learning* yang mengelompokkan data ke dalam sejumlah *cluster* berdasarkan kemiripan fitur. Dalam konteks pengolahan citra, fitur yang sering digunakan adalah tekstur, *Gray Level Co-occurrence Matrix* (*GLCM*) digunakan untuk memperoleh fitur-fitur tekstur dari citra. [5]

Proses segmentasi citra motif batik berdasarkan tekstur dimulai dari tahap “*input citra*” ke sistem, kemudian dilakukan “*preprocessing*” berupa konversi citra berwarna menjadi grayscale, normalisasi ukuran, dan pengurangan noise untuk mempersiapkan data. Selanjutnya, dilakukan “ekstraksi fitur tekstur menggunakan metode *GLCM* (*Gray Level Co-occurrence Matrix*)” dengan menghitung hubungan antar piksel pada jarak dan sudut tertentu, menghasilkan

parameter seperti *contrast*, *correlation*, *energy*, dan *homogeneity* yang mewakili karakteristik tekstur motif batik. Vektor fitur yang dihasilkan dari *GLCM* kemudian menjadi “*input* bagi algoritma *K-Means Clustering*” untuk proses pengelompokan. *K-Means* akan menginisialisasi *centroid* secara acak, menghitung jarak *Euclidean* antara setiap vektor fitur dengan *centroid*, menetapkan setiap data ke *cluster* terdekat, lalu memperbarui posisi *centroid* hingga *konvergen*. Hasil akhirnya adalah pengelompokan citra batik ke dalam *cluster* berdasarkan kemiripan tekstur, yang dapat digunakan untuk analisis pola, katalogisasi, atau pengembangan sistem pengenalan motif batik.

Berbagai penelitian terkini menunjukkan keunggulan *K-Means Clustering*. Pemilihan metode *k-means* didasarkan pada sifatnya sebagai metode analisis data tanpa pengawasan (*unsupervised*) yang melakukan pengelompokan data dengan sistem partisi. *Clustering* merupakan proses pengelompokan data atau objek ke dalam *cluster* (grup) sehingga objek dalam satu *cluster* memiliki kemiripan yang tinggi dan berbeda dari objek di *cluster* lain.

Berdasarkan Penelitian terdahulu yang dilakukan oleh Yasmin, Nabilla Chairunnissa, Syifa, Akbar, Deliva, Ramadhanu, Agung “Penerapan *K-Means Clustering* untuk Klasifikasi Citra Cabai Keriting : Studi Ekstraksi Warna dan Tekstur *GLCM*”. penelitian menunjukkan bahwa penerapan metode *K-Means Clustering* yang dikombinasikan dengan ekstraksi fitur warna menggunakan model *RGB* dan *Lab*, serta tekstur menggunakan metode *Grey Level Co-occurrence Matrix (GLCM)*, mampu mengklasifikasikan citra cabai keriting merah dan hijau dengan tingkat akurasi yang tinggi. Dari 100 citra yang

digunakan, masing-masing kelas (merah dan hijau) diuji dengan 50 citra dan diperoleh akurasi sebesar 93% untuk cabai merah dan 91% untuk cabai hijau, dengan rata-rata keseluruhan 92%. Hasil ini membuktikan bahwa integrasi antara ciri warna dan tekstur secara efektif meningkatkan performa klasifikasi citra dan dapat diterapkan dalam sistem otomatisasi pertanian.[6]

Penelitian terdahulu yang dilakukan oleh Agyztia Premana, Raden Mohamad, Herdian Bhakti, Dimas Prayogi “Segmentasi *K-Means Clustering* pada citra menggunakan Ekstraksi Fitur Warna Dan Tekstur”. penelitian menunjukkan bahwa metode segmentasi citra menggunakan ekstraksi fitur warna dan tekstur dengan *filter Gabor* dan algoritma *K-Means Clustering* mampu meningkatkan akurasi segmentasi citra digital. *Filter Gabor* digunakan untuk mengekstraksi tekstur, sementara ruang warna *Lab* digunakan untuk ekstraksi warna. Meskipun *filter Gabor* memiliki kelemahan pada citra dengan banyak tekstur makro, integrasi metode *K-Means* berhasil memperbaiki akurasi. Penelitian mencatat peningkatan penggunaan fitur tekstur hingga 17,5% dan warna keabu-abuan sebesar 16,24%, dengan tambahan akurasi dari filter Gabor sebesar 2% dan peningkatan 0,3% pada ruang warna *Lab*, menunjukkan efektivitas pendekatan gabungan ini dalam segmentasi citra.[2]

Berdasarkan Penelitian terdahulu yang dilakukan oleh Supriyatna, Samsu “Klasifikasi Data Menggunakan Algoritma *K-Means Clustering* Dan *Naive Bayes Classifier* Berdasarkan Analisa Tekstur Metode *Gray Level Co-Occurrence Matrix (GLCM)*”. penelitian menunjukkan bahwa penggunaan algoritma *K-Means Clustering* dan *Naive Bayes Classifier* dalam klasifikasi data citra digital

berdasarkan analisis tekstur dengan metode *Gray Level Co-occurrence Matrix* (*GLCM*) menghasilkan akurasi yang sangat baik, yaitu 100% untuk kedua algoritma. Penelitian ini menggunakan 66 sampel citra digital dengan format .gif, dan fitur tekstur yang digunakan adalah *contrast* dan *correlation* pada sudut 0°, 45°, 90°, dan 135°, serta jarak piksel -1. *GLCM* terbukti efektif dalam mengekstraksi ciri tekstur dari citra digital, sementara *K-Means* dan *Naive Bayes* berhasil mengelompokkan data ke dalam tiga cluster secara akurat berdasarkan karakteristik tersebut.[7]

Hasil *clustering* dari algoritma *K-means* mudah di interpretasi, memungkinkan identifikasi pola dan tren dalam segmentasi motif batik berdasarkan tekstur. Hal ini memudahkan analisis lebih lanjut dan pengambilan kesimpulan yang relevan. Algoritma *K-means* dipilih karena efisiensi komputasinya yang tinggi, kemampuannya dalam mengelompokkan data berdasarkan kemiripan fitur tekstur, serta kemudahan interpretasi hasil. Karakteristik ini sangat penting untuk segmentasi motif batik yang efektif dan efisien.

Dengan semua kelebihan tersebut, penelitian ini bertujuan untuk merancang dan mengimplementasikan sistem segmentasi motif batik secara otomatis berdasarkan tekstur menggunakan algoritma *K-Means Clustering*. Implementasi ini diharapkan dapat menjadi solusi *efisien* untuk mengklasifikasikan dan mengidentifikasi motif batik secara cepat dan akurat. Sistem ini dapat membantu dalam berbagai aplikasi, seperti identifikasi motif batik asli, pengarsipan motif batik, pengembangan desain batik baru, dan pengawasan kualitas produksi batik.

Selain itu, penelitian ini juga berkontribusi pada pengembangan teknologi pengolahan citra digital dan analisis pola pada kain batik.

1.2 Rumusan Masalah

Rumusan masalah yang diangkat dalam penelitian ini adalah bagaimana merancang dan mengimplementasikan Algoritma *K-means Clustering* Untuk Segmentasi Citra Motif Batik Berdasarkan Tekstur?

1.3 Tujuan Masalah

Tujuan dari penelitian ini adalah Merancang dan mengimplementasikan Algoritma *K-means Clustering* Untuk Segmentasi Citra Motif Batik Berdasarkan Tekstur.

1.4 Batasan Masalah

Batasan masalah dalam penelitian ini adalah :

1. Data yang digunakan berupa citra batik dengan 1600 Data kaggle dengan nama DataSet “Kumpulan Data Batik Nusantara (Batik Indonesia)”.
2. Segmentasi motif batik hanya dilakukan berdasarkan tekstur, tanpa mempertimbangkan faktor warna atau pola motif lainnya.
3. Algoritma *K-Means Clustering* yang digunakan tidak dibandingkan dengan algoritma segmentasi citra lainnya. Pra-pemrosesan citra dilakukan dengan teknik standar pengolahan citra digital, *Gray Level Co-occurrence Matrix (GLCM)*.

1.5 Manfaat Penelitian

Penelitian ini memiliki manfaat yang signifikan dalam meningkatkan pemahaman dan penerapan algoritma *K-Means Clustering* dalam pengolahan citra, terutama untuk segmentasi motif batik berdasarkan tekstur. Dengan menggunakan fitur-fitur tekstur seperti energi, kontras, korelasi, dan homogenitas, penelitian ini memberikan kontribusi pada peningkatan efektivitas dalam mengelompokkan motif batik yang kompleks dan beragam. Selain itu, Penelitian ini memberikan kontribusi dalam pengembangan ilmu pengetahuan, khususnya di bidang pengolahan citra digital dan analisis data. Dengan menerapkan algoritma *K-Means* untuk segmentasi motif batik berdasarkan tekstur, penelitian ini membantu menunjukkan bagaimana metode tersebut bisa digunakan untuk mengenali pola-pola yang rumit dalam gambar batik. Selain memberikan pemahaman baru tentang cara kerja *K-Means* dalam konteks citra budaya lokal, hasil penelitian ini juga bisa menjadi dasar untuk pengembangan sistem lain, seperti sistem rekomendasi desain batik, menemukan cara membuat motif batik yang baru dan membuka jalan bagi penelitian-penelitian selanjutnya di bidang yang sama. Dengan demikian, segmentasi citra batik berdasarkan tekstur tidak hanya memberikan manfaat praktis bagi industri, tetapi juga berkontribusi pada pelestarian dan pengembangan budaya batik.

1.6 Sistematika Penulisan

Sistematika penulisan dari skripsi ini terdiri dari pokok-pokok permasalahan yang akan dibahas pada masing-masing yang diuraikan menjadi beberapa bagian sebagai berikut :

BAB 1 PENDAHULUAN

Bab ini berisi latar belakang, rumusan masalah, tujuan penelitian, batasan masalah, manfaat penelitian, dan sistematika penulisan.

BAB 2 LANDASAN TEORI

Bab ini berisi teori-teori yang mendasari penelitian, meliputi pengolahan citra digital, ekstraksi fitur tekstur (*Gray Level Co-occurrence Matrix GLCM*, *Wavelet Transform*), algoritma *clustering*, dan algoritma *K-Means Clustering*. Penjelasan mengenai konsep segmentasi citra dan penerapannya pada pengolahan citra batik juga akan diuraikan.

BAB 3 METODOLOGI PENELITIAN

Bab ini menjelaskan tahapan penelitian, mulai dari pengumpulan data citra batik, pra-pemrosesan citra (konversi ke *grayscale*, *noise reduction*), ekstraksi fitur tekstur, implementasi algoritma *K-Means Clustering*, penentuan nilai *K* (jumlah *cluster*), evaluasi hasil segmentasi.

BAB 4 ANALISIS DAN PERANCANGAN SISTEM

Bab ini berisi analisis kebutuhan sistem untuk segmentasi motif batik, perancangan sistem yang meliputi alur proses segmentasi, pemilihan parameter algoritma *K-Means*.

BAB 5 IMPLEMENTASI DAN PENGUJIAN

Bab ini menjelaskan implementasi sistem segmentasi motif batik berbasis *K-Means Clustering*, hasil pengujian sistem terhadap dataset citra batik, dan analisis hasil segmentasi yang meliputi evaluasi kinerja sistem berdasarkan metrik yang telah ditentukan.

BAB 6 PENUTUP

Bab ini berisi kesimpulan dari penelitian yang telah dilakukan serta saran untuk penelitian atau pengembangan sistem di masa mendatang.

BAB 2

LANDASAN TEORI

2.1 *Artificial Intelligence (AI)*

Artificial Intelligence (AI) atau biasa disebut dengan kecerdasan buatan merupakan sesuatu yang memungkinkan mesin untuk dapat melakukan pekerjaan yang dapat dilakukan oleh manusia. Kecerdasan ini memerlukan pengetahuan, pengalaman dan kemampuan dalam mengambil keputusan dan melakukan tindakan.[8] *Artificial Intelligence (AI)* adalah cabang ilmu komputer yang bertujuan untuk mengembangkan sistem dan mesin yang mampu melakukan tugas yang biasanya memerlukan kecerdasan manusia.

AI melibatkan penggunaan algoritma dan model matematika untuk memungkinkan komputer dan sistem lainnya untuk belajar dari data, mengenali pola, dan membuat keputusan yang cerdas. Dalam konteks *AI*, terdapat beberapa konsep penting seperti *machine learning* (pembelajaran mesin), *neural networks* (jaringan saraf tiruan), *natural language processing* (pemrosesan bahasa alami), dan banyak lagi. Pengembangan *AI* telah memberikan dampak besar dalam berbagai bidang seperti pengenalan suara, pengenalan wajah, mobil otonom, pengobatan, dan masih banyak lagi.[9]

2.2 *Deep Learning*

Deep learning adalah sebuah teknologi yang mampu bekerja menggunakan beberapa algoritma tertentu. Sejatinya, tidak ada algoritma *deep learning* yang dianggap sempurna. Sebab, masing-masing jenis memiliki

kapabilitas yang berbeda. Maka dari itu, application developer harus memilih jenis algoritma yang paling sesuai dengan kebutuhan mereka. Untuk memilih algoritma yang tepat, ada baiknya kamu pahami masing-masing jenis algoritma *deep learning*. [10]

Deep learning adalah cabang dari machine learning yang secara khusus fokus pada penggunaan *neural networks* (jaringan saraf tiruan) dengan banyak lapisan untuk meniru cara kerja otak manusia dalam memproses informasi. Istilah "*deep*" merujuk pada kedalaman lapisan dalam jaringan saraf tersebut, di mana model terdiri dari beberapa lapisan yang secara bertahap mampu memahami pola yang semakin kompleks dan abstrak dalam data. Teknologi ini menjadi dasar dalam banyak inovasi di bidang kecerdasan buatan *Artificial Intelligence (AI)*, seperti pengenalan suara, pemrosesan bahasa alami, visi komputer, dan lain sebagainya. [11]

Deep learning adalah salah satu inovasi terbesar dalam bidang kecerdasan buatan yang telah membawa dampak signifikan di berbagai sektor, mulai dari kesehatan, keuangan, hingga industri kreatif. Teknologi ini merupakan cabang dari *machine learning* yang memungkinkan mesin untuk belajar dan mengenali pola-pola kompleks melalui jaringan saraf tiruan yang terinspirasi dari cara kerja otak manusia. Dengan struktur yang terdiri dari lapisan-lapisan neuron, *deep learning* mampu memproses data dalam skala besar, memungkinkan analisis yang lebih mendalam dan akurasi yang lebih tinggi dibandingkan metode-metode tradisional. [11]

2.3 *Machine Learning*

Machine learning merupakan cabang ilmu bagian dari kecerdasan buatan (*artificial intelligence*), dengan pemrograman untuk memungkinkan komputer menjadi cerdas berperilaku seperti manusia, dan dapat meningkatkan pemahamannya melalui pengalaman. secara otomatis. *Machine learning* memiliki fokus pada pengembangan sistem yang mampu belajar sendiri untuk memutuskan sesuatu tanpa harus berulang kali diprogram oleh manusia. Hal ini menjadikan mesin tidak hanya mampu berperilaku mengambil keputusan, namun juga dapat beradaptasi dengan perubahan yang terjadi. *Machine learning* bekerja apabila tersedia data sebagai input untuk dilakukan analisis terhadap kumpulan data besar (*big data*) sehingga menemukan pola tertentu. Data merupakan bahan input yang akan digunakan untuk melakukan pembelajaran (*training*) sehingga mesin dapat menghasilkan analisis yang benar. Didalam *machine learning* dikenal *data training* dan *data testing*, *data training* untuk melatih algoritma dalam *machine learning* dan *data testing* untuk mengetahui performa dari algoritma dalam *machine learning* yang telah dilatih yaitu ketika menemukan data baru yang belum pernah diberikan dalam *data training*. [12]

Machine learning menggunakan teknik untuk menangani data besar (*big data*) dengan cara yang cerdas untuk memberikan hasil yang tepat. Berdasarkan teknik pembelajarannya, tipe-tipe *machine learning* dapat dibedakan menjadi *supervised learning*, *unsupervised learning*, *semi supervised learning* dan *reinforcement learning*. *Supervised learning* merupakan salah satu teknik *machine learning* yang menggunakan dataset (*data training*) yang sudah berlabel (*labeled*

data) untuk melakukan pembelajaran pada mesin, sehingga mesin mampu mengidentifikasi label input dengan menggunakan fitur yang dimiliki untuk selanjutnya melakukan prediksi maupun klasifikasi, sedangkan *unsupervised learning* adalah teknik dengan menarik kesimpulan berdasarkan dataset yang merupakan input data *labeled response*. [12]

2.4 *Data Mining*

Data mining atau penambangan data adalah proses ekstraksi informasi yang berguna dari kumpulan data besar. Teknik ini menggabungkan statistik dan kecerdasan buatan untuk menganalisis data dan menemukan pola yang berguna. *Data mining* dapat membantu perusahaan dalam mengubah data mentah menjadi pengetahuan yang berguna. Teknik-teknik *data mining* yang mendasari analisis ini dapat dibagi menjadi dua tujuan utama; mereka dapat menggambarkan kumpulan data target atau mereka dapat memprediksi hasil melalui penggunaan algoritma pembelajaran mesin. *Data mining* dapat disebut sebagai suatu proses pencarian atau penambangan demi mendapatkan informasi dari sekumpulan data yang berjumlah besar (*Big data*). [13]

Metode-metode ini digunakan untuk mengorganisir dan menyaring data, menampilkan informasi yang paling menarik, dari deteksi penipuan hingga perilaku pengguna, bottleneck, dan bahkan pelanggaran keamanan. Ketika digabungkan dengan alat analisis data dan visualisasi, seperti *Apache Spark*, menjelajahi dunia *data mining* belum pernah semudah ini dan mengekstrak wawasan yang relevan belum pernah secepat ini. Kemajuan dalam kecerdasan buatan terus mempercepat adopsi di seluruh industri. Proses data mining

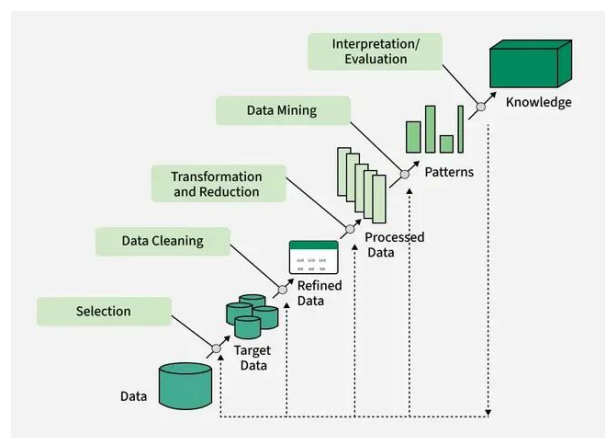
melibatkan sejumlah langkah mulai dari pengumpulan data hingga visualisasi untuk mengekstrak informasi yang berguna dari kumpulan data besar. Teknik teknik *data mining* digunakan untuk menghasilkan deskripsi dan prediksi tentang kumpulan data target.[9]

Data mining dapat disebut sebagai suatu proses pencarian atau penambangan demi mendapatkan informasi dari sekumpulan data yang berjumlah besar (*Big data*). *Data mining* sendiri sering disebut sebagai *knowledge discovery in database (KDD)* adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola hubungan dalam set data berukuran besar. *Output* dari *data mining* ini dapat digunakan untuk pengambilan keputusan di masa depan. Salah satu teknik yang dikenal dalam *data mining* yaitu *clustering*. Berupa proses pengelompokan sejumlah data atau objek ke dalam *cluster* (group) sehingga setiap dalam *cluster* tersebut akan berisi data yang semirip mungkin dan berbeda dengan objek dalam cluster yang lainnya.

Penemuan Pengetahuan dalam *Basis Data (KDD)* mengacu pada proses lengkap untuk mengungkap pengetahuan berharga dari kumpulan data besar. Dimulai dengan pemilihan data yang relevan, diikuti dengan praproses untuk membersihkan dan mengaturnya, transformasi untuk mempersiapkannya untuk analisis, penambangan data untuk mengungkap pola dan hubungan, dan diakhiri dengan evaluasi dan interpretasi hasil, yang pada akhirnya menghasilkan pengetahuan atau wawasan yang berharga. *KDD* banyak digunakan dalam bidang-bidang seperti pembelajaran mesin, pengenalan pola, statistik, kecerdasan buatan, dan visualisasi data. *Data mining* bertujuan untuk mengekstraksi esensi

pengetahuan dari sekumpulan data sehingga dapat disajikan dalam bentuk struktur yang dapat dipahami manusia. Dalam proses penemuan pengetahuan dalam database (*Knowledge Discovery in Database – KDD*).[14]

Proses *KDD* bersifat iteratif, yang melibatkan penyempurnaan berulang untuk memastikan keakuratan dan keandalan pengetahuan yang diekstraksi. Seluruh proses terdiri dari langkah-langkah berikut:



Gambar 2. 1 Proses Data Mining KDD (Knowledge Discovery in Database)

- a. Seleksi Data (*Selection*)
- b. Pemilihan data (*Preprocessing / Cleaning*)
- c. Transformasi (*Transformation*)
- d. *Data Mining*
- e. Interpretasi / Evaluasi (*Interpretation / Evaluation*)

2.5 Algoritma *K-Means Clustering*

Algoritma *K-means clustering* merupakan metode paling populer digunakan untuk mendapatkan deskripsi dari sekumpulan data dengan cara mengungkapkan kecenderungan setiap data lainnya. Kecenderungan pengelompokkan tersebut didasarkan kemiripan karakteristik individu-individu

data yang ada. Ide dasar dari teknik ini adalah menemukan pusat dari setiap kelompok data yang mungkin ada untuk kemudian mengelompokkan pada setiap data individu kedalam salah satu dari kelompok-kelompok tersebut berdasarkan jaraknya.[2]

- a. Untuk rumus normalisasi *Min-Max*, yang digunakan untuk mengubah skala data agar berada dalam rentang tertentu, biasanya $[0, 1]$. Ini sangat umum digunakan dalam *data mining* dan *machine learning*, termasuk saat menggunakan algoritma *K-Means Clustering*.

$$X_{norm} = \frac{(X - X_{min})}{(X_{max} - X_{min})} \dots \dots \dots (1)$$

- 1) X adalah Nilai asli (data mentah sebelum dinormalisasi)
- 2) X_{min} adalah Nilai minimum dari seluruh data fitur tersebut
- 3) X_{max} adalah Nilai maksimum dari seluruh data fitur tersebut
- 4) X_{norm} adalah Nilai hasil normalisasi (nilai setelah diskalakan)

- b. rumus jarak *Euclidean*, yang digunakan untuk menghitung jarak antara dua titik dalam ruang berdimensi banyak (multidimensional), seperti dalam algoritma *K-Means Clustering*.

$$d = \sqrt{\sum (x_i - c_i)^2} \dots \dots \dots (2)$$

di mana:

- 1) d adalah Jarak *Euclidean* antara satu data (x) dan satu centroid (c).
- 2) x_i adalah Nilai fitur ke- i dari titik data (misalnya *Energy*, *Contrast*, dll).
- 3) c_i adalah Nilai fitur ke- i dari *centroid*.

4) $\sum$ adalah Penjumlahan terhadap seluruh fitur yang dibandingkan

Rumus ini digunakan untuk:

- a) Menentukan seberapa dekat suatu titik data ke pusat *cluster* (*centroid*)
- b) Mengelompokkan data ke *cluster* dengan jarak paling kecil (terdekat)

K-Means Clustering adalah salah satu teknik *clustering* yang paling umum digunakan dalam analisis data. Algoritma ini bekerja dengan cara membagi data ke dalam K kelompok yang telah ditentukan sebelumnya, di mana K merupakan jumlah kelompok yang diinginkan. Proses dimulai dengan memilih secara acak K titik pusat kelompok (*centroid*) di dalam ruang data, lalu mengelompokkan setiap titik data ke dalam kelompok yang memiliki *centroid* terdekat. Kemudian, titik pusat setiap kelompok dihitung kembali berdasarkan rata-rata titik data dalam kelompok tersebut, dan proses ini diulangi hingga tidak ada lagi perubahan dalam penempatan titik data ke dalam kelompok atau hingga batasan literasi yang ditentukan tercapai. *K-Means Clustering* sering digunakan dalam segmentasi pelanggan, pengelompokan dokumen, analisis citra, dan bidang lainnya karena kecepatan dan skalabilitasnya, meskipun hasilnya dapat dipengaruhi oleh *inisialisasi awal centroid*.

Selain kecepatan dan skalabilitasnya, *K-Means Clustering* juga memiliki kemampuan untuk menangani data yang besar dengan baik, membuatnya cocok untuk analisis pada dataset yang besar seperti yang sering ditemui dalam aplikasi bisnis dan ilmiah. Namun, perlu dicatat bahwa keberhasilan *K-Means* sangat

tergantung pada pemilihan jumlah kelompok yang tepat (nilai K) dan inisialisasi awal centroid yang baik.[15]

K-Means merupakan salah satu algoritma clustering yang digunakan dalam proses *Data mining*. *K-Means* bekerja dengan cara melakukan pengelompokan secara partisi yang memisahkan data kedalam kelompok-kelompok tertentu dengan meminimalkan rata-rata jarak setiap data klasternya.[13]

Clustering adalah sebuah metode pengelompokan data berdasarkan informasi yang didapatkan dari data yang menguraikan objek tersebut. Pada metode *clustering* terdapat beberapa algoritma salah satunya adalah algoritma *K-Means Clustering*. *K-Means Clustering* merupakan salah satu teknik pengelompokan data dengan cara membagi data yang ada kedalam bentuk satu atau lebih *cluster* Karena kesederhanaan memiliki Tingkat ketelitian yang tinggi terhadap ukuran data sehingga *K-Means Clustering* lebih efisien dan terukur dalam pengolahan data berjumlah besar, selain itu urutan objek tidak mempengaruhi algoritma ini.[16]

2.6 Pengolahan Citra Digital

Pengolahan citra digital adalah suatu pengolahan citra yang bertujuan meningkatkan kualitas citra agar lebih mudah dipahami oleh manusia dan komputer. Dalam proses pengolahan citra digital terdapat 2 proses yaitu input dan output. Hasil dari pengolahan citra digital berupa objek dalam bentuk citra atau gambar. Pengolahan citra digital mempunyai kelebihan seperti tidak merusak objek, pengolahannya cepat dan mudah. Citra adalah data dalam format gambar

dan biasanya memiliki ukuran citra yang besar, semakin bagus citra yang dihasilkan maka kapasitas citra yang dimiliki semakin besar pula. Di era *digitalisasi*, banyak orang memberikan informasi melalui gambar karena informasi dengan menggunakan gambar lebih mudah dipahami dan efektif.

Citra dapat dikelompokkan menjadi 2 jenis yaitu citra *analog* dan citra *digital*. Citra *analog* merupakan citra yang terbentuk dari sinyal analog bersifat kontinu. contoh citra analog seperti foto sinar-X, gambar yang dihasilkan oleh monitor televisi dan lain-lain. Sedangkan citra *digital* adalah citra yang dihasilkan melalui proses digitalisasi citra *analog* contoh citra *digital* seperti hasil foto dari kamera digital dan *scanner*. Citra *digital* adalah gambar 2 dimensi yang dibuat dari *analog* 2 dimensi yang kontinu menjadi citra melalui proses *sampling*. Citra *analog* dibagi menjadi N baris dan M kolom, menjadikannya gambar diskrit. Citra *digital* adalah citra yang dapat diproses dan disimpan oleh komputer, citra yang terdapat di dalam komputer hanyalah angka yang menunjukkan intensitas pada setiap piksel. Karena berbentuk data numerik, maka citra digital dapat diproses oleh dengan komputer.[17]

2.7 Motif Batik dan Karakteristiknya

Batik merupakan sebuah karya seni yang berasal dari budaya Indonesia dengan corak yang beragam dengan mengadaptasi berbagai bentuk dari eksplorasi alam maupun kebudayaan yang ada di Indonesia. Serta tata warna yang khas dari setiap daerah yang menunjukkan identitas bangsa Indonesia. Sejak tanggal 2 Oktober 2009, *UNESCO* salah satu badan *PBB* yang menangani bagian pendidikan dan kebudayaan telah mengukuhkan bahwa batik adalah warisan

budaya dunia tak benda yang berasal asli dari Indonesia, karena dalam hal ini yang diwariskan bukanlah benda batiknya, akan tetapi ilmunya maka dari itu batik disebut sebagai warisan budaya dunia tak benda.[18]

Motif batik adalah pola atau ragam hias yang menjadi elemen utama dalam kain batik, yang dibuat melalui teknik pewarnaan khusus menggunakan malam (lilin). Motif-motif ini tidak hanya berfungsi sebagai ornamen, tetapi juga mengandung makna simbolik dan nilai budaya yang mencerminkan filosofi, sejarah, serta identitas suatu daerah atau kelompok masyarakat di Indonesia. Motif-motif batik yang dikembangkan oleh para pengrajin mempunyai makna yang tersendiri, sehingga dalam setiap motif yang ada tidak bisa disamakan maknanya dengan batik yang lain.[19]

Menurut Kementerian Pendidikan dan Kebudayaan Republik Indonesia, batik merupakan warisan budaya tak benda yang mengandung unsur seni, kerajinan, dan teknologi tradisional. Setiap motif batik biasanya memiliki nama, makna, serta aturan penggunaan tertentu, yang sering dikaitkan dengan status sosial, acara adat, atau nilai-nilai kehidupan.

Karakteristik motif batik terletak pada pola hiasnya yang bersifat repetitif, simetris, dan kaya akan makna simbolik yang mencerminkan nilai-nilai budaya, filosofi hidup, serta identitas kedaerahan. Setiap motif biasanya memiliki struktur visual yang unik, berupa kombinasi bentuk geometris maupun organik seperti garis, lengkung, flora, dan fauna, yang disusun dengan kerapatan tertentu sehingga membentuk tekstur khas. Selain itu, motif batik juga berbeda-beda antar daerah di Indonesia misalnya Batik Yogyakarta cenderung klasik dan geometris,

sementara Batik Riau cenderung memiliki tekstur halus dan pola longgar, sehingga pendekatan segmentasi perlu mengandalkan fitur tekstur lembut dan pewarnaan gradasi untuk membedakan motif utama dari latar.. Tekstur visual yang dihasilkan dari susunan motif ini menjadi salah satu ciri penting dalam pengenalan dan segmentasi citra batik secara digital.[20]

2.8 Fitur Tekstur *GLCM (Gray Level Co-occurrence Matrix)*

Tekstur dalam citra digital merupakan pola *visual* yang merepresentasikan struktur permukaan suatu objek dalam gambar, berdasarkan distribusi intensitas piksel yang membentuk kesan *visual* seperti kasar, halus, berulang, atau acak. Tekstur mencerminkan keteraturan atau ketidakteraturan susunan elemen visual, seperti garis, titik, atau warna, yang membentuk pola lokal pada suatu permukaan.

Metode yang digunakan untuk mengekstraksi fitur tekstur dari citra digital ialah *GLCM (Gray Level Co-occurrence Matrix)* menganalisis hubungan antara piksel berdasarkan jarak dan arah. *GLCM* adalah suatu matriks yang elemennya merupakan jumlah pasangan piksel yang memiliki tingkat kecerahan tertentu, di mana pasangan piksel itu terpisah dengan jarak d , dan dengan suatu sudut inklinasi θ . Dengan kata lain, *matriks kookurensi* adalah *probabilitas* munculnya *gray level* i dan j dari dua piksel yang terpisah pada jarak d dan sudut θ . [21]

2.9 Segmentasi Citra

Segmentasi citra adalah proses dalam pengolahan citra digital yang bertujuan untuk membagi citra menjadi beberapa bagian (segmen) yang lebih

bermakna dan lebih mudah dianalisis. Setiap segmen biasanya merepresentasikan suatu objek atau area tertentu yang memiliki karakteristik seragam, seperti warna, tekstur, intensitas, atau pola. Segmentasi digunakan untuk memisahkan objek utama dari latar belakang atau untuk mengelompokkan bagian-bagian gambar berdasarkan kemiripan fitur visual. Proses ini merupakan salah satu tahapan penting dalam sistem pengenalan pola, analisis citra medis, pengawasan video, dan pengenalan motif seperti batik.

Segmentasi adalah proses pembagian citra menjadi beberapa daerah atau objek. Dalam segmentasi citra, terdapat pendekatan *discontinuity* dan *similarity* berdasarkan intensitas piksel. Pendekatan *discontinuity* membagi citra berdasarkan perubahan intensitas yang tiba-tiba (berdasarkan tepi). Pendekatan *similarity* membagi citra menjadi daerah-daerah yang memiliki sifat yang serupa (berdasarkan wilayah), seperti *thresholding*, *region growing*, *region splitting*, dan *merging*. Tujuan dari segmentasi citra adalah memisahkan wilayah objek dari latar belakang agar objek dapat dianalisis dengan mudah dalam konteks pengenalan objek yang melibatkan persepsi *visual*. [22]

2.10 *Kaggle*

Kaggle adalah platform komunitas daring yang dirancang untuk mendukung kegiatan data science dan *machine learning*. Platform ini menyediakan berbagai sumber daya seperti forum diskusi, notebook interaktif, kompetisi, dan yang paling relevan untuk penelitian ini, yaitu kumpulan dataset publik yang dapat digunakan secara gratis. Dataset yang tersedia di *Kaggle* mencakup berbagai bidang seperti pengolahan citra, teks, prediksi numerik, dan

banyak lainnya.

Dalam konteks pengolahan citra digital, *Kaggle* menjadi sumber terpercaya untuk memperoleh dataset citra berkualitas tinggi, termasuk anotasi dan metadata yang sering kali dibutuhkan dalam tahap preprocessing dan pelabelan. Salah satu keunggulan utama *Kaggle* adalah keterbukaannya terhadap komunitas *global*, sehingga data yang tersedia berasal dari berbagai institusi, proyek *open-source*, maupun hasil kontribusi pengguna individu.

Dalam penelitian ini, *Kaggle* digunakan sebagai sumber data untuk citra motif batik yang akan diproses menggunakan algoritma *K-Means Clustering*. Pemanfaatan data dari *Kaggle* mendukung prinsip *reproducibility* dan validasi karena *dataset* yang digunakan bersifat publik dan dapat diakses oleh siapa pun secara bebas. Data yang digunakan berupa citra batik dengan 1600 berkas(Data *Kaggle*).

Data yang digunakan merupakan data sekunder yang diunduh dari *www.kaggle.com*, *platform* kompetisi data science menggunakan perintah pengunduhan melalui *API Kaggle*. Data yang telah diunduh kemudian akan dianalisis menggunakan *software Python* dengan *library Scikit-learn* dan *XGBoost*. [23]

2.11 *Python*

Python adalah bahasa pemrograman dinamis, tingkat tinggi, dimana merupakan bahasa pemrograman interpreter yaitu bahasa yang mengkonversi *source code* menjadi *machine code* secara langsung ketika program dijalankan. Bahasa ini juga mendukung pendekatan pemrograman Berorientasi Objek untuk

pengembangan aplikasi dan mudah dipelajari serta menyediakan banyak struktur data tingkat tinggi. *Python* adalah bahasa skrip yang mudah dipelajari namun kuat dan serbaguna, yang membuatnya menarik untuk Pengembangan Aplikasi. *Sintaks* dan pengetikan *Python* sangat dinamis dengan sifat interpretasinya menjadikannya bahasa yang ideal untuk skrip dan pengembangan aplikasi yang cepat. *Python* mendukung banyak pola pemrograman, termasuk gaya pemrograman berorientasi objek, imperatif, dan fungsional serta prosedural.[24]

Python mendukung multi paradigma pemrograman, utamanya, namun tidak dibatasi pada pemrograman berorientasi objek, pemrograman imperatif, dan pemrograman fungsional. Salah satu fitur yang tersedia pada *python* adalah sebagai bahasa pemrograman dinamis yang dilengkapi dengan manajemen memori otomatis. Seperti halnya pada bahasa pemrograman dinamis lainnya, *python* umumnya digunakan sebagai bahasa skrip meski pada praktiknya penggunaan bahasa ini lebih luas mencakup konteks pemanfaatan yang umumnya tidak dilakukan dengan menggunakan bahasa script. *Python* dapat digunakan untuk berbagai keperluan pengembangan perangkat lunak dan dapat berjalan di berbagai platform sistem operasi, Saat ini kode python dapat dijalankan di berbagai platform sistem operasi.[25] Berbagai library python digunakan dalam penelitian ini antara lain:

1. **OpenCV (Open Source Computer Vision)**

OpenCV adalah perpustakaan berbasis "Sumber Terbuka" dengan lebih dari 500 fungsi untuk menangani visi komputer. Perangkat lunak ini dapat digunakan untuk tujuan bisnis dan komersial dan didistribusikan

di bawah lisensi BSD. Platform seperti *Windows*, *Linux*, *Mac OS*, *iOS*, dan *Android* didukung. *Open Source Computer Vision* adalah tempat asal nama *OpenCV*. Bahasa C digunakan untuk membuat *OpenCV*. Python dapat digunakan sebagai antarmuka untuk mengakses perpustakaan *OpenCV*. Dapat diunduh di <https://opencv.org>. [26]

2. *Numpy (Numerical Python)*

Numerical Python adalah pustaka fundamental untuk komputasi numerik di *Python*. Dikenalkan pada tahun 2005, *NumPy* menyediakan dukungan untuk array multidimensi dan berbagai fungsi matematika dan statistik yang efisien. [27]

3. *TensorFlow*

TensorFlow adalah kerangka kerja komputasi untuk membangun model pembelajaran mesin. *TensorFlow* menyediakan berbagai toolkit yang memungkinkan untuk membuat model pada tingkat abstraksi yang disukai dan dapat menjalankan grafik pada beberapa platform hardware, termasuk *CPU*, *GPU*, dan *TPU*. [28]

4. *Keras*

Keras merupakan suatu *package* yang berfungsi sebagai *wrapper* *TensorFlow* untuk lebih memudahkan user dalam penggunaannya. Namun keras bukan hanya bisa untuk *TensorFlow* namun bisa digunakan untuk yang lainnya seperti *Theano* dan *CNTK*. [29]

5. Pandas

Pandas adalah pustaka Python yang dirancang untuk manipulasi dan analisis data. Diperkenalkan pada tahun 2008 oleh Wes McKinney, Pandas menawarkan struktur data yang fleksibel dan alat analisis yang kuat, memungkinkan pengguna untuk bekerja dengan data terstruktur dengan mudah.[27]

2.12 *Google Colab (Colaboratory)*

Google Colab merupakan perangkat yang sangat berguna bagi pemula dan pakar di bidang *data science*, *machine learning*, dan penanganan data secara umum. *Google Colab* telah menjadi pilihan populer untuk banyak proyek dan tugas komputasi berbasis *Python* karena sumber daya komputasi yang luas, kemudahan berbagi dan lingkungan yang interaktif.[30]

Google Collaboratory, adalah editor kode berbasis *cloud* yang tersedia secara gratis untuk tujuan penelitian. Dibangun dengan elemen -elemen dari *jupyter*, *google colab* menyediakan hampir semua library yang diperlukan untuk penelitian, termasuk *data mining*. [30]

2.13 *Anaconda Navigator*

Anaconda Navigator adalah *graphical user interface* atau biasa dikenal sebagai *GUI* yang berfungsi untuk menjalankan aplikasi dan mengelola paket yang nantinya digunakan untuk memanggil *library* dalam kode program yang dibutuhkan oleh pengguna. Didalam *anaconda navigator* terdapat beberapa aplikasi salah satunya adalah *jupyter*. [31]

2.14 *Flowchart*

Flowchart adalah mengembangkan aturan logika dari suatu prosedur pemecahan masalah, sehingga *flowchart* merupakan langkah-langkah penyelesaian masalah yang dituliskan dalam simbol tertentu.[32]

Flowchart adalah penggambaran secara grafik dari langkah-langkah dan urutan prosedur dari suatu program. *Flowchart* sistem merupakan suatu urutan proses dalam system dengan menunjukkan alat dari media input, output serta jenis media yang digunakan untuk penyimpanan dalam proses pengolahan data sedangkan *flowchart* program merupakan suatu bagan dengan simbol-simbol tertentu yang menggambarkan suatu urutan dari proses secara detail dan berhubungan antara suatu proses (instruksi) dengan proses lainnya dalam suatu program . [33]

2.15 *Unified Model Language (UML)*

Unified Model Language (UML) merupakan suatu himpunan struktur dan teknik untuk pemodelan dan desain *object oriented programming (OOP)* beserta aplikasinya. *UML* adalah suatu metodologi dalam mengembangkan sistem *OOP* dan sekelompok *tool* untuk mendukung pengembangan sistem tersebut.[34]

1. *Use Case Diagram*

Use Case Diagram adalah diagram yang menggambarkan interaksi antar sistem dengan system eksternal dan pengguna. Diagram Use Case digunakan untuk mengetahui fungsi apa saja yang ada di dalam sebuah sistem dan siapa saja yang berhak menggunakan fungsi–fungsi tersebut.[34]

2. *Activity Diagram*

diagram aktivitas (*activity diagram*) menggambarkan *workflow* (aliran kerja) atau aktivitas dari sebuah sistem atau proses bisnis. *Activity diagram* menggambarkan urutan kegiatan dimana diagram ini mendukung tingkah laku kondisional dan paralel. *Activity diagram* memodelkan fungsi – fungsi suatu sistem dan aliran kendali antar objek.[34]

3. *Class Diagram*

diagram kelas atau class diagram menggambarkan struktur sistem dari segi pendefinisian kelas - kelas yang akan dibuat untuk membangun sistem. Kelas memiliki apa yang disebut atribut dan metode atau operasi, yaitu : Atribut merupakan variabel-variabel yang dimiliki oleh suatu kelas dan operasi atau metode adalah fungsi yang dimiliki oleh suatu kelas. *Class diagram* dibuat untuk menyediakan sebuah gambaran atau pandangan dari beberapa atau semua class dalam sebuah model. *Class diagram* mendeskripsikan jenis-jenis objek dalam sistem dan berbagai macam hubungan interaksi di dalamnya.[34]

4. *Sequence Diagram*

Sequence Diagram merupakan salah satu interaction diagram yang menjelaskan bagaimana suatu operasi itu dilakukan; message (pesan) apa yang dikirim dan kapan pelaksanaannya. Diagram ini diatur berdasarkan waktu. Objek - objek yang berkaitan dengan proses berjalannya operasi diurutkan dari kiri ke kanan berdasarkan waktu terjadinya dalam pesan yang terurut. Diagram sekuen (*sequence diagram*) menggambarkan kelakuan objek pada *Use Case* dengan

mendeskripsikan waktu hidup objek dan message yang dikirimkan dan diterima antar objek

2.16 Penelitian Terkait

Berikut adalah penelitian sebelumnya yang dijadikan acuan dan referensi dalam menyelesaikan masalah ini.

Tabel 2. 1 Tabel Penelitian Sebelumnya

No	Nama	Tahun	Judul	Hasil
1	Muhibbul, Muhammad	2023	Segmentasi Citra Daun Bawang Merah Menggunakan Metode Thresholding Otsu	Hasil penelitian menunjukkan bahwa metode Thresholding Otsu mampu melakukan segmentasi citra daun bawang merah secara efektif dengan akurasi yang tinggi. Pengujian dilakukan terhadap 20 citra dengan berbagai kondisi pencahayaan dan latar belakang melalui dua skenario utama. Skenario terbaik diperoleh pada citra dengan latar belakang pasir gelap dan pencahayaan redup, menghasilkan rata-rata nilai PSNR sebesar 50,42 dB dan MSE sebesar 0,74622, yang menunjukkan kualitas segmentasi yang sangat baik. Secara keseluruhan, rata-rata PSNR dari seluruh skenario mencapai 49,34744 dB dan MSE 0,85681, membuktikan bahwa metode ini mampu menghasilkan segmentasi yang akurat dan cocok diterapkan pada pengolahan citra pertanian, khususnya

				untuk objek seperti daun bawang merah.[22]
2.	Ula, Mutammimul Zulfikri, Abdi Ulva, Ananda Faridhatul Rizal, Reyhan Achmad	2023	Penerapan Machine Learning Clustering K-Means dan Linear Regression dalam Penentuan Tingkat Resiko Tuberkulosis Paru	Hasil penelitian menunjukkan bahwa algoritma <i>K-Means Clustering</i> berhasil mengelompokkan 17 kecamatan di Kabupaten Bireuen ke dalam tiga klaster tingkat risiko tuberkulosis paru, yaitu 2 daerah masuk klaster 1, 6 daerah dalam klaster 2, dan 9 daerah dalam klaster 3. Sementara itu, penerapan <i>Clusterwise Regression</i> menghasilkan nilai koefisien determinasi (R^2) sebesar 0,5740, yang mengindikasikan bahwa sekitar 57% variasi jumlah pasien TB paru dapat dijelaskan oleh variabel kepadatan penduduk. Hasil ini menunjukkan bahwa kombinasi algoritma K-Means dan regresi linier efektif untuk memetakan daerah rawan serta memprediksi pengaruh faktor demografis terhadap penyebaran TB paru, sehingga dapat dijadikan referensi bagi dinas kesehatan dalam penanganan penyakit ini.[35]
3.	Hendrastuty, Nirwana	2024	Penerapan Data Mining Menggunakan Algoritma K-Means Clustering dalam Evaluasi Hasil Pembelajaran Siswa	Hasil penelitian menunjukkan bahwa penerapan algoritma K-Means Clustering berhasil mengelompokkan data hasil pembelajaran siswa ke dalam dua klaster, yaitu kelompok siswa Rajin (C0) sebanyak 63 siswa dan kelompok siswa Sangat

				Rajin (C1) sebanyak 91 siswa. Proses clustering dilakukan berdasarkan nilai akademik dari berbagai mata pelajaran, yang sebelumnya telah melalui tahap pembersihan dan normalisasi data. Evaluasi menggunakan silhouette score menghasilkan nilai sebesar 0,9168, yang mengindikasikan bahwa pengelompokan yang dihasilkan sangat baik dan optimal. Hasil ini memberikan gambaran bahwa K-Means efektif digunakan untuk mengidentifikasi pola pencapaian belajar siswa serta dapat dimanfaatkan sebagai dasar dalam penyusunan strategi pembelajaran yang lebih tepat sasaran.[15]
4.	Adawiyah, Quratih Defit, Sarjon	2024	Penerapan Algoritma K-Means Clustering untuk Mengelompokkan Rekomendasi Metode Kontrasepsi Berbasis Machine Learning	Hasil penelitian menunjukkan bahwa algoritma K-Means Clustering efektif digunakan untuk mengelompokkan data pasien KB menjadi dua klaster, yaitu klaster C0 untuk metode kontrasepsi jangka pendek dan klaster C1 untuk metode jangka panjang. Dari 20 data pasien yang digunakan, proses clustering menghasilkan distribusi akhir sebanyak 11 anggota di klaster C0 dan 9 anggota di klaster C1, dengan proses iterasi yang stabil pada iterasi ke-4. Penggunaan aplikasi RapidMiner membantu dalam pemrosesan dan

				visualisasi hasil, serta mempercepat proses analisis data. Hasil akhir membuktikan bahwa algoritma ini mampu mengidentifikasi pola dalam data pasien berdasarkan usia dan jumlah anak, sehingga dapat digunakan untuk mendukung pemberian rekomendasi metode kontrasepsi yang lebih akurat dan efisien oleh tenaga kesehatan di Puskesmas.
5.	Kanugroho, Muhammad Tegar Rahman, Muh Arif Wihandika, Randy Cahya	2022	Klasifikasi Batik dengan Ekstraksi Fitur Tekstur Local Binary Pattern dan Metode K-Nearest Neighbor	Hasil penelitian menunjukkan bahwa ekstraksi fitur tekstur menggunakan metode Local Binary Pattern (LBP) yang mencakup mean, standard deviation, dan skewness mampu membedakan kelas motif batik secara efektif. Proses klasifikasi dilakukan menggunakan algoritma K-Nearest Neighbor (KNN) dengan berbagai nilai tetangga (k), dan hasil terbaik diperoleh pada nilai $k = 5$, dengan akurasi tertinggi mencapai 65% setelah proses normalisasi data. Pengujian dilakukan dengan beberapa kombinasi fitur dan random state, yang menunjukkan bahwa kombinasi fitur lengkap (mean, standard deviation, dan skewness) memberikan performa klasifikasi yang paling stabil. Temuan ini membuktikan bahwa

				kombinasi metode LBP dan KNN dapat digunakan sebagai pendekatan awal dalam sistem klasifikasi citra batik berbasis tekstur.[36]
6.	Yasmin, Nabilla Chairunnissa, Syifa Akbar, Deliva Ramadhanu, Agung	2024	Penerapan <i>K-Means Clustering</i> untuk Klasifikasi Citra Cabai Keriting: Studi Ekstraksi Warna dan Tekstur GLCM	Hasil penelitian menunjukkan bahwa algoritma <i>K-Means Clustering</i> efektif dalam mengklasifikasikan citra cabai keriting merah dan hijau dengan mengombinasikan ekstraksi fitur warna (RGB dan Lab) serta tekstur (GLCM). Proses klasifikasi dilakukan pada dataset yang terdiri dari 100 citra, masing-masing 50 citra untuk setiap kelas, dengan pembagian 60% untuk pelatihan dan 40% untuk pengujian. Pengujian akurasi menunjukkan hasil yang sangat baik, yaitu 93% untuk citra cabai merah keriting dan 91% untuk cabai hijau keriting, dengan rata-rata akurasi keseluruhan sebesar 92%. Hasil ini membuktikan bahwa kombinasi fitur warna dan tekstur, bersama dengan metode segmentasi berbasis K-Means, mampu menangkap karakteristik visual cabai keriting secara efektif dan dapat diterapkan dalam sistem otomatisasi klasifikasi objek pertanian.[6]

7.	Yulian, Dwi RL Panca Budiarto, Sony	2022	Penerapan Algoritma K- Means Clustering untuk Segmentasi Pelanggan B- Clean Laundry Banyuwangi	Penelitian ini berhasil Hasil penelitian menunjukkan bahwa algoritma <i>K-Means Clustering</i> berhasil digunakan untuk melakukan segmentasi pelanggan B-Clean Laundry berdasarkan data transaksi bulan Mei 2022. Dari total 442 transaksi dan 240 pelanggan, dilakukan normalisasi data menggunakan metode Min-Max dan pengelompokan ke dalam dua hingga empat klaster. Evaluasi menggunakan rasio BCV/WCV menunjukkan bahwa jumlah klaster terbaik adalah 2, dengan nilai rasio terkecil sebesar 0,009, menandakan kualitas klaster yang tinggi. Hasil pengelompokan membagi pelanggan ke dalam dua kelompok, yaitu 237 pelanggan sebagai pelanggan reguler dan 3 pelanggan sebagai pelanggan premium. Segmentasi ini dapat dimanfaatkan untuk strategi pemasaran yang lebih tepat sasaran, seperti pemberian reward atau promo khusus, guna meningkatkan loyalitas dan pendapatan usaha.[37]
8.	Muhammad, Imam Fakhri Hendrik, Billy Putra, Universitas Yptk, Indonesia	2025	Penerapan Algoritma K- Means Clustering untuk Segmentasi	Hasil penelitian menunjukkan bahwa algoritma <i>K-Means Clustering</i> memiliki potensi tinggi dalam

			Berdasarkan Kesiapan Seleksi: Systematic Literature Review	mengelompokkan individu berdasarkan tingkat kesiapan seleksi, khususnya dalam konteks pendidikan dan pelatihan. Kajian sistematis terhadap 15 artikel ilmiah membuktikan bahwa K-Means mampu membentuk klaster yang jelas dan bermakna berdasarkan berbagai parameter kesiapan seperti nilai akademik, hasil tryout, kehadiran, dan akses terhadap sumber belajar. Segmentasi ini memberikan gambaran objektif yang dapat dimanfaatkan oleh lembaga pendidikan atau bimbingan belajar untuk menyusun strategi pembelajaran yang lebih adaptif dan personal. Meskipun demikian, penerapan K-Means juga menghadapi tantangan seperti sensitivitas terhadap outlier dan penentuan jumlah klaster optimal. Secara keseluruhan, penelitian ini menegaskan bahwa K-Means merupakan pendekatan inovatif dan efisien dalam mendukung proses seleksi yang berbasis data dan lebih tepat sasaran.[38]
9.	Efran, Fernando Ade Pratama Khairil Jumadi, Juju	2022	Implementasi Metode K-Means Clustering pada Segmentasi Citra Digital	Hasil penelitian menunjukkan bahwa algoritma <i>K-Means Clustering</i> berhasil diterapkan untuk segmentasi dan identifikasi objek pada citra digital dengan hasil yang cukup akurat, khususnya pada objek dengan warna khas atau

				dominan. Pengujian dilakukan menggunakan beberapa citra uji bebas seperti laptop, bunga tulip, pisang, strawberry, dan sepeda, yang menghasilkan tingkat kemiripan antara 74,45% hingga 96,73% terhadap data pelatihan. Sistem dikembangkan menggunakan bahasa pemrograman <i>Visual Basic .NET</i>, dan segmentasi dilakukan berdasarkan kedekatan nilai warna piksel. Proses clustering berhasil mengelompokkan piksel dengan karakteristik serupa ke dalam klaster yang sama, sehingga memudahkan dalam proses identifikasi objek secara otomatis. Penelitian ini membuktikan bahwa metode K-Means efektif dalam pengelompokan warna dan dapat digunakan sebagai dasar untuk sistem pengenalan objek dalam citra digital.[39]
10.	Rosiani, Ulla Delfana Rahmad, Cahya Rahmawati, Marcelina Alifia Tupamahu, Frangky	2020	Segmentasi Berbasis K-Means pada Deteksi Citra Penyakit Daun Tanaman Jagung	Hasil penelitian menunjukkan bahwa metode <i>K-Means Clustering</i> dengan pendekatan warna menggunakan transformasi RGB ke CIE Lab* dan Euclidean Distance efektif dalam mendeteksi jenis penyakit daun jagung, yaitu hawar daun dan bercak daun. Penelitian dilakukan dengan 30 citra untuk pelatihan dan 10 citra untuk pengujian. Hasil segmentasi citra

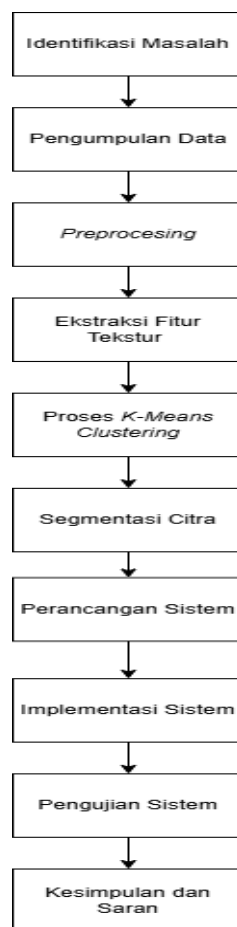
				menunjukkan bahwa sistem mampu mengidentifikasi citra uji sesuai dengan data pelatihan dengan tingkat akurasi sebesar 90%. Proses segmentasi menggunakan nilai pixel dari histogram a^* untuk membentuk tiga kluster ($k=3$), dan hasil pengujian membuktikan bahwa pendekatan ini berhasil memisahkan objek berdasarkan perbedaan warna yang signifikan. Sistem juga telah diuji secara fungsional melalui antarmuka berbasis Python dan Django, yang berjalan sesuai harapan dalam keseluruhan proses mulai dari input gambar hingga deteksi hasil penyakit.[40]
--	--	--	--	---

BAB 3

METODOLOGI PENELITIAN

3.1 Kerangka Penelitian

Penelitian ini dilakukan secara bertahap dengan pendekatan yang terstruktur dan sistematis. Setiap tahapan saling berkaitan untuk memastikan proses penelitian berjalan dengan baik dan menghasilkan model klasifikasi yang bagus. Metode yang digunakan digambarkan secara rinci dalam skema yang menggambarkan alur penelitian secara jelas dan teratur. Adapun tahapan lengkap dari penelitian ini dapat dilihat pada Gambar 3.1 berikut:



Gambar 3. 1 Tahapan Penelitian

3.1.1 Identifikasi Masalah

Dalam pengolahan citra *digital* batik, salah satu tantangan utama adalah bagaimana memisahkan motif dari latar belakang atau elemen lain dalam gambar. Proses ini dikenal dengan segmentasi citra, yang bertujuan membagi gambar ke dalam beberapa bagian yang memiliki karakteristik serupa. Motif batik yang kaya akan variasi tekstur dan kompleksitas pola sering kali sulit diproses menggunakan metode segmentasi *konvensional* berbasis warna atau bentuk. Oleh karena itu, diperlukan pendekatan segmentasi yang mempertimbangkan ciri tekstur untuk menghasilkan pemisahan motif yang lebih akurat. Dalam hal ini, algoritma *K-Means Clustering* menjadi salah satu metode yang potensial untuk diterapkan karena mampu mengelompokkan data secara otomatis berdasarkan kemiripan fitur. Berdasarkan hal tersebut, masalah yang diidentifikasi dalam penelitian ini adalah bagaimana mengimplementasikan algoritma *K-Means Clustering* untuk segmentasi citra motif batik dengan memperhatikan fitur tekstur secara efektif dan efisien.

3.1.2 Pengumpulan Data

Tahap ini berfokus pada pengumpulan citra motif batik yang akan digunakan dalam penelitian. Data berupa gambar batik dengan variasi motif dan tekstur yang beragam dikumpulkan untuk memastikan representasi yang komprehensif. Sumber data diambil dari www.kaggle.com dengan *Data Kaggle* sebanyak 1600 berkas gambar dengan nama DataSet “Kumpulan Data Batik Nusantara (Batik Indonesia)”. Kualitas gambar perlu diperhatikan untuk mendapatkan hasil segmentasi yang optimal.

3.1.3 *Preprocessing*

Data citra mentah yang dikumpulkan biasanya belum siap untuk diproses lebih lanjut. Tahap *Preprocessing* bertujuan untuk membersihkan dan mempersiapkan data agar sesuai dengan kebutuhan algoritma *K-Means*. *Preprocessing* adalah proses menyiapkan data mentah untuk dianalisis atau digunakan dalam model pembelajaran mesin. Ini adalah langkah penting dalam proses pembelajaran mesin, karena kualitas dan karakteristik data dapat memengaruhi performa model secara signifikan. Proses tahap processing dilakukan guna pengolahan data mentah menjadi set data yang sudah jadi untuk memilih data yang akan diproses dalam dokumen.[41] Proses ini dapat meliputi beberapa langkah, seperti:

- a. *Resize*: Mengubah ukuran gambar agar seragam dan efisien dalam komputasi.
- b. *Noise Reduction*: Mengurangi noise atau gangguan pada gambar untuk meningkatkan kualitas data.
- c. *Enhancement*: Meningkatkan kontras atau ketajaman gambar untuk menonjolkan fitur tekstur.

3.1.4 **Ekstraksi Fitur Tekstur**

Tahap ini merupakan inti dari segmentasi berbasis tekstur. Fitur-fitur tekstur diekstrak dari citra batik yang telah dipraproses. Beberapa metode ekstraksi fitur tekstur yang umum digunakan antara lain:

- a. *GLCM (Gray-Level Co-occurrence Matrix)*: Menghitung statistik tekstur berdasarkan kemunculan pasangan piksel dengan nilai intensitas tertentu.

- b. *Haralick Features*: Seperangkat fitur tekstur yang dihitung dari *GLCM* (*Gray-Level Co-occurrence Matrix*), seperti energi, kontras, korelasi, dll.
- c. *Wavelet Transform*: Menggunakan transformasi *wavelet* untuk menganalisis tekstur pada berbagai skala resolusi.

Fitur-fitur yang diekstrak akan digunakan sebagai input untuk algoritma *K-Means*.

3.1.5 Proses *K-Means Clustering*

Algoritma *K-Means* digunakan untuk mengelompokkan piksel-piksel dalam citra batik berdasarkan kesamaan fitur tekstur yang telah diekstrak. Algoritma ini akan membagi data menjadi beberapa kelompok (*cluster*) yang berbeda, dimana setiap cluster mewakili suatu jenis tekstur pada motif batik. Jumlah cluster (*K*) perlu ditentukan terlebih dahulu, dan pemilihan nilai *K* yang tepat sangat berpengaruh pada hasil segmentasi.

K-Means merupakan salah satu metode data clustering non-hirarki yang mengelompokkan data dalam bentuk satu atau lebih cluster/kelompok. *K-Means* merupakan algoritma clustering dengan metode partisi (*partitioning method*) yang berbasis titik pusat (*centroid*) selain algoritma *k-Medoids* yang berbasis objek.[42]

3.1.6 Segmentasi Citra

Setelah proses *clustering* selesai, setiap piksel dalam citra batik akan diberi label sesuai dengan cluster yang ditempatinya. Hasilnya adalah citra tersegmentasi dimana setiap region mewakili suatu jenis tekstur. Tahap ini

menghasilkan visualisasi segmentasi tekstur pada motif batik.

Teknik segmentasi citra sebagai bagian dari proses pengolahan citra, adalah kegiatan untuk membagi citra menjadi beberapa bagian atau region, yang bertujuan untuk mengisolasi atau menemukan suatu obyek di dalam citra. Tujuan segmentasi citra adalah untuk mempartisi gambar menjadi beberapa wilayah yang tidak tumpang tindih dengan karakteristik yang homogen, seperti intensitas, warna, dan tekstur. Selain hal tersebut segmentasi citra juga bertujuan terhadap objek dengan proses segmentasi masing-masing objek pada citra dapat diambil secara terpisah sehingga dapat digunakan sebagai masukan proses yang lain dan mempermudah untuk dikembangkan oleh proses selanjutnya.[43]

3.1.7 Perancangan Sistem

Tahap ini meliputi perancangan sistem atau aplikasi yang akan mengimplementasikan algoritma dan proses yang telah dijelaskan sebelumnya. Perancangan ini mencakup aspek antarmuka pengguna, alur proses, dan pemilihan teknologi yang tepat.

3.1.8 Implementasi Sistem

Beberapa komponen pendukung yang berperan penting dalam implementasi sistem diantaranya adalah perangkat keras (*hardware*) dan perangkat lunak (*software*). Adapun spesifikasi dari perangkat keras dan perangkat lunak yang digunakan sebagai berikut:

1. Perangkat keras (*hardware*), antara lain:

Processor : Intel Core i5

Memory : 8GB RAM

System type : *Windows 11 Home Single Language 64-bit (10.0, Build 26100)*

Harddisk : 500 GB

2. Perangkat lunak (*software*), antara lain:

Sistem operasi : *Windows 11*

Tools : *Jupyter Notebook, Anaconda_Prompt*

3. *Platform Eksperimen: Google Colaboratory, Anaconda Navigator, Streamlit, Web Browser*

3.1.9 Pengujian Sistem

Pengujian sistem merupakan tahapan penting dalam pengembangan perangkat lunak, termasuk sistem segmentasi citra motif batik ini. Tujuan utama dari pengujian sistem adalah untuk memastikan bahwa seluruh komponen sistem bekerja sesuai dengan yang diharapkan dan memenuhi kebutuhan fungsional serta non-fungsional yang telah ditetapkan. Pengujian ini melibatkan serangkaian aktivitas yang dirancang untuk mengidentifikasi potensi kesalahan (*bugs*), mengevaluasi kinerja sistem, serta memvalidasi bahwa sistem telah memenuhi standar kualitas yang diinginkan. Proses pengujian sistem ini dilakukan setelah tahap implementasi sistem selesai.

Dalam konteks penelitian ini, pengujian sistem dilakukan untuk memverifikasi keefektifan algoritma *K-Means Clustering* dan metode ekstraksi fitur tekstur *GLCM* dalam melakukan segmentasi citra motif batik. Pengujian ini mencakup pengujian fungsionalitas, seperti memastikan bahwa sistem dapat

menerima input citra batik, melakukan pra-pemrosesan dengan benar, mengekstraksi fitur tekstur yang relevan, menjalankan algoritma *K-Means Clustering* dengan tepat, dan menghasilkan output segmentasi yang akurat. Selain itu, pengujian juga dilakukan untuk mengevaluasi kinerja sistem dari segi kecepatan pemrosesan, penggunaan sumber daya, serta kemampuan menangani berbagai variasi citra batik dengan kompleksitas tekstur yang berbeda-beda. Hasil dari pengujian sistem ini akan digunakan sebagai dasar untuk mengevaluasi keberhasilan penelitian dan memberikan rekomendasi perbaikan atau pengembangan lebih lanjut.

Sistem yang telah dibuat diuji untuk mengevaluasi keakuratan dan performa segmentasi dapat dilakukan dengan *Blackbox*. Pengujian *blackbox* dilakukan untuk memastikan bahwa setiap fitur dalam sistem berfungsi sebagaimana mestinya dan sesuai dengan tujuan yang telah ditetapkan. Pengujian ini bertujuan untuk mengidentifikasi apakah masih terdapat kesalahan dalam program atau apakah sistem telah berjalan dengan baik dan sesuai harapan.

3.1.10 Kesimpulan dan Saran

Tahap terakhir menyimpulkan hasil penelitian, menganalisis kelebihan dan kekurangan metode yang digunakan, serta memberikan saran untuk penelitian selanjutnya. Kesimpulan harus didukung oleh data dan hasil pengujian yang telah dilakukan.