



Hak Cipta Dilindungi Undang-Undang

1. Dilarang memperbanyak atau mendistribusikan dokumen ini untuk tujuan komersial tanpa izin tertulis dari penulis atau pihak berwenang.
2. Penggunaan tanpa izin untuk kepentingan akademik, penelitian, dan pendidikan diperbolehkan dengan mencantumkan sumber.
3. Plagiarisme juga dilarang dan dapat dikenakan sanksi.
3. Universitas hanya berhak menyimpan dan mendistribusikan dokumen ini di repositori akademik, tanpa mengalihkan hak cipta penulis, sesuai dengan peraturan yang berlaku di Indonesia.

Universitas Islam Indragiri

BAB III

METODE PENELITIAN

Metode penelitian yang digunakan dalam penelitian ini adalah penelitian kuantitatif dengan pendekatan data mining menggunakan algoritma K-Means *clustering*. Pendekatan kuantitatif dipilih karena penelitian ini menganalisis data numerik berupa jumlah penjualan (qty) dan total penjualan (total) per produk. Data penjualan yang digunakan merupakan data historis Toko Azhar Tani Mandiri periode tahun 2024–2025 yang diperoleh secara langsung dari toko. Penelitian ini tidak hanya berfokus pada proses analisis data, tetapi juga menghasilkan sistem berbasis PWA yang mampu menjalankan proses *clustering* secara otomatis dan menampilkan hasil pengelompokan produk dalam bentuk visualisasi. Sistem dibangun untuk membantu pemilik Toko Azhar Tani Mandiri dalam mengetahui produk mana yang paling berdampak pada tingkat penjualan toko, sehingga keputusan pembelian stok dapat dilakukan berdasarkan data penjualan yang terukur.

3.1 Kerangka Penelitian

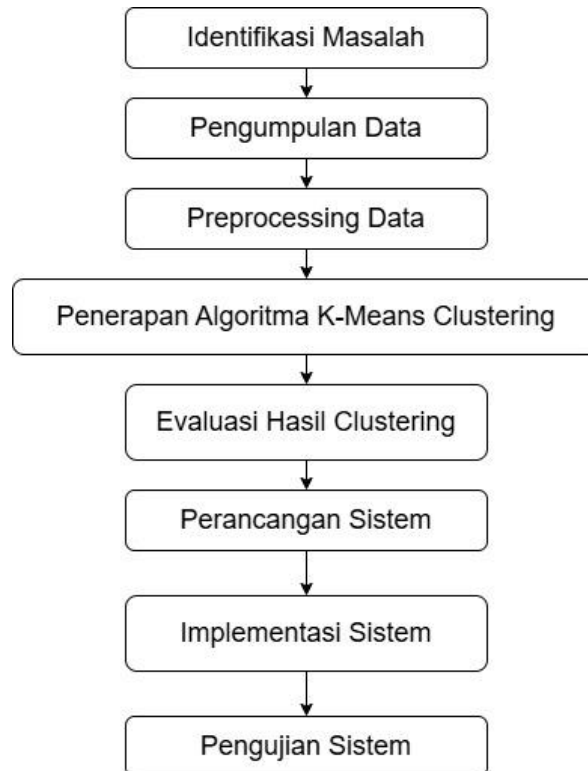
Kerangka penelitian disusun untuk memberikan gambaran yang jelas mengenai alur pelaksanaan penelitian dari tahap awal hingga tahap akhir. Kerangka ini berfungsi sebagai pedoman agar penelitian berjalan secara sistematis, terarah, dan konsisten dengan tujuan yang ingin dicapai. Adapun kerangka penelitian yang dibuat seperti Gambar 3.1 dibawah ini:



Hak Cipta Dilindungi Undang-Undang

1. Dilarang memperbanyak atau mendistribusikan dokumen ini untuk tujuan komersial tanpa izin tertulis dari penulis atau pihak berwenang.
2. Penggunaan tanpa izin untuk kepentingan akademik, penelitian, dan pendidikan diperbolehkan dengan mencantumkan sumber.
3. Universitas hanya berhak menyimpan dan mendistribusikan dokumen ini di repositori akademik, tanpa mengalihkan hak cipta penulis, sesuai dengan peraturan yang berlaku di Indonesia.

Universitas Islam Indragiri



Gambar 3.1 Kerangka Penelitian

Gambar 3.1 penelitian ini memiliki tahapan – tahapan sistematis sebagai berikut:

3.1.1 Identifikasi Masalah

Tahap identifikasi masalah dilakukan untuk mengetahui permasalahan yang terjadi pada Toko Azhar Tani Mandiri. Berdasarkan hasil observasi awal dan wawancara dengan pemilik toko, diketahui bahwa data penjualan sudah tercatat secara otomatis melalui sistem pencatatan transaksi harian yang telah tersedia, namun data tersebut belum diolah lebih lanjut untuk menghasilkan informasi pengelompokan produk berdasarkan tingkat penjualan toko. Penentuan jumlah pembelian ulang produk masih dilakukan berdasarkan perkiraan dan pengalaman tanpa didukung analisis data yang sistematis. Kondisi tersebut menyebabkan pemilik toko belum dapat mengetahui secara terukur produk mana yang paling



1. Dilarang memperbanyak atau mendistribusikan dokumen ini untuk tujuan komersial tanpa izin tertulis dari penulis atau pihak berwenang.
 2. Penggunaan tanpa izin untuk kepentingan akademik, penelitian, dan pendidikan diperbolehkan dengan mencantumkan sumber.
 3. Plagiarisme juga dilarang dan dapat dikenakan sanksi.
 3. Universitas hanya berhak menyimpan dan mendistribusikan dokumen ini di repositori akademik, tanpa mengalihkan hak cipta penulis, sesuai dengan peraturan yang berlaku di Indonesia.

berdampak pada tingkat penjualan, mana yang cukup diminati, dan mana yang

jarang dibeli pelanggan. Akibatnya, pembelian stok belum dilakukan secara tepat sasaran, sehingga diperlukan pendekatan berbasis data untuk mendukung keputusan pembelian produk.

3.1.2 Pengumpulan Data

Tahap pengumpulan data bertujuan untuk memperoleh data yang dibutuhkan dalam penelitian. Data yang digunakan terdiri dari data primer dan data sekunder sebagai berikut:

3.1.1.1 Data Primer

Data primer diperoleh secara langsung dari Toko Azhar Tani Mandiri melalui:

- Observasi, dilakukan dengan mengamati secara langsung proses operasional Toko Azhar Tani Mandiri, khususnya pada proses pencatatan data transaksi penjualan produk, guna mengidentifikasi permasalahan yang dihadapi dalam menentukan performa penjualan tiap produk.
- Wawancara, dilakukan secara langsung dengan pemilik Toko Azhar Tani Mandiri pada tanggal 22 Januari 2026, yang bertujuan untuk memperoleh informasi mengenai proses pencatatan transaksi penjualan, kendala yang dihadapi dalam mengelompokkan atau menentukan performa penjualan tiap produk.
- Dokumentasi, dilakukan dengan mengumpulkan data penjualan produk Toko Azhar Tani Mandiri periode tahun 2024–2025. Data yang dikumpulkan meliputi nama barang, tanggal transaksi, jumlah penjualan (qty), harga



Hak Cipta Dilindungi Undang-Undang

1. Dilarang memperbanyak atau mendistribusikan dokumen ini untuk tujuan komersial tanpa izin tertulis dari penulis atau pihak berwenang.
 2. Penggunaan tanpa izin untuk kepentingan akademik, penelitian, dan pendidikan diperbolehkan dengan mencantumkan sumber.
 3. Universitas hanya berhak menyimpan dan mendistribusikan dokumen ini di repositori akademik, tanpa mengalihkan hak cipta penulis, sesuai dengan peraturan yang berlaku di Indonesia.

Universitas Islam Indragiri

satuan, dan total penjualan (total), yang selanjutnya digunakan sebagai dataset dalam proses *clustering*.

3.1.1.2 Data Sekunder

Data sekunder dalam penelitian ini diperoleh melalui studi literatur yang berkaitan dengan topik penelitian. Studi literatur ini seperti jurnal ilmiah dan artikel penelitian terdahulu yang relevan dengan penelitian ini.

3.1.3 Preprocessing Data

Preprocessing dilakukan untuk meningkatkan kualitas data sebelum dimasukkan ke proses analisis. Tahapan yang dilakukan meliputi:

- Seleksi data dilakukan dengan menentukan atribut yang digunakan dalam proses *clustering*, yaitu jumlah penjualan (qty) dan total penjualan (total). Atribut nama produk digunakan sebagai identitas data untuk memudahkan interpretasi hasil *clustering*.
- Pembersihan data dilakukan untuk memastikan data yang digunakan tidak mengandung nilai kosong, data duplikat, atau kesalahan pencatatan yang dapat memengaruhi hasil *clustering*.
- Normalisasi data menggunakan metode Min-Max, untuk menyamakan skala nilai antar atribut sehingga perhitungan jarak antar data menjadi lebih akurat. Algoritma K-Means sensitif terhadap skala data, maka digunakan normalisasi, misalnya Min-Max Normalization, dengan rumus:

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (1)$$

3.1.4 Penerapan Algoritma K-Means Clustering

Algoritma K-Means digunakan untuk mengelompokkan produk berdasarkan jumlah penjualan (qty) dan total penjualan (total) menjadi tiga *cluster*, yaitu produk



Hak Cipta Dilindungi Undang-Undang

1. Dilarang memperbanyak atau mendistribusikan dokumen ini untuk tujuan komersial tanpa izin tertulis dari penulis atau pihak berwenang.
Penggunaan untuk kepentingan akademik, penelitian, dan pendidikan diperbolehkan dengan mencantumkan sumber.

2. Penggunaan tanpa izin untuk kepentingan komersial atau pelanggaran hak cipta dapat dikenakan sanksi sesuai dengan UU Hak Cipta di Indonesia. Plagiarisme juga dilarang dan dapat dikenakan sanksi.

3. Universitas hanya berhak menyimpan dan mendistribusikan dokumen ini di repositori akademik, tanpa mengalihkan hak cipta penulis, sesuai dengan peraturan yang berlaku di Indonesia.

Hak Cipta Dilindungi Undang-Undang

Universitas Islam Indragiri

paling laris, produk cukup laris, dan produk kurang laris. Tahapan penerapan algoritma K-Means adalah sebagai berikut:

1. Menentukan jumlah *cluster* (k), yaitu $k = 3$.
2. Menentukan centroid awal secara acak.
3. Menghitung jarak setiap data penjualan ke masing-masing centroid menggunakan metode *Euclidean Distance*.

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \quad (2)$$

Dimana :

$d(x,y)$ = jarak *Euclidean* antara data x dan centroid y

n = jumlah atribut (dalam penelitian ini $n = 2$, yaitu qty dan total)

x_i = atribut dari data ke- i , ($i = 1, 2, 3, \dots, n$)

y_i = atribut dari pusat *Cluster* ke- i , ($i = 1, 2, 3, \dots, n$)

4. Mengelompokkan data ke dalam *cluster* dengan jarak terdekat ke centroid.
5. Menghitung ulang centroid berdasarkan rata-rata data pada setiap *cluster*.

$$C_k = \frac{(a_1 + a_2 + a_3 + \dots + a_n)}{\Sigma a} \quad (3)$$

Dimana :

C_k = centroid *cluster* ke k

a_1 = nilai record data ke 1

a_2 = nilai record data ke 2

Σa = jumlah record data

6. Melakukan iterasi hingga tidak terjadi perubahan centroid atau hasil pengelompokan.

3.1.4.1 Penentuan Jumlah *Cluster* Optimal



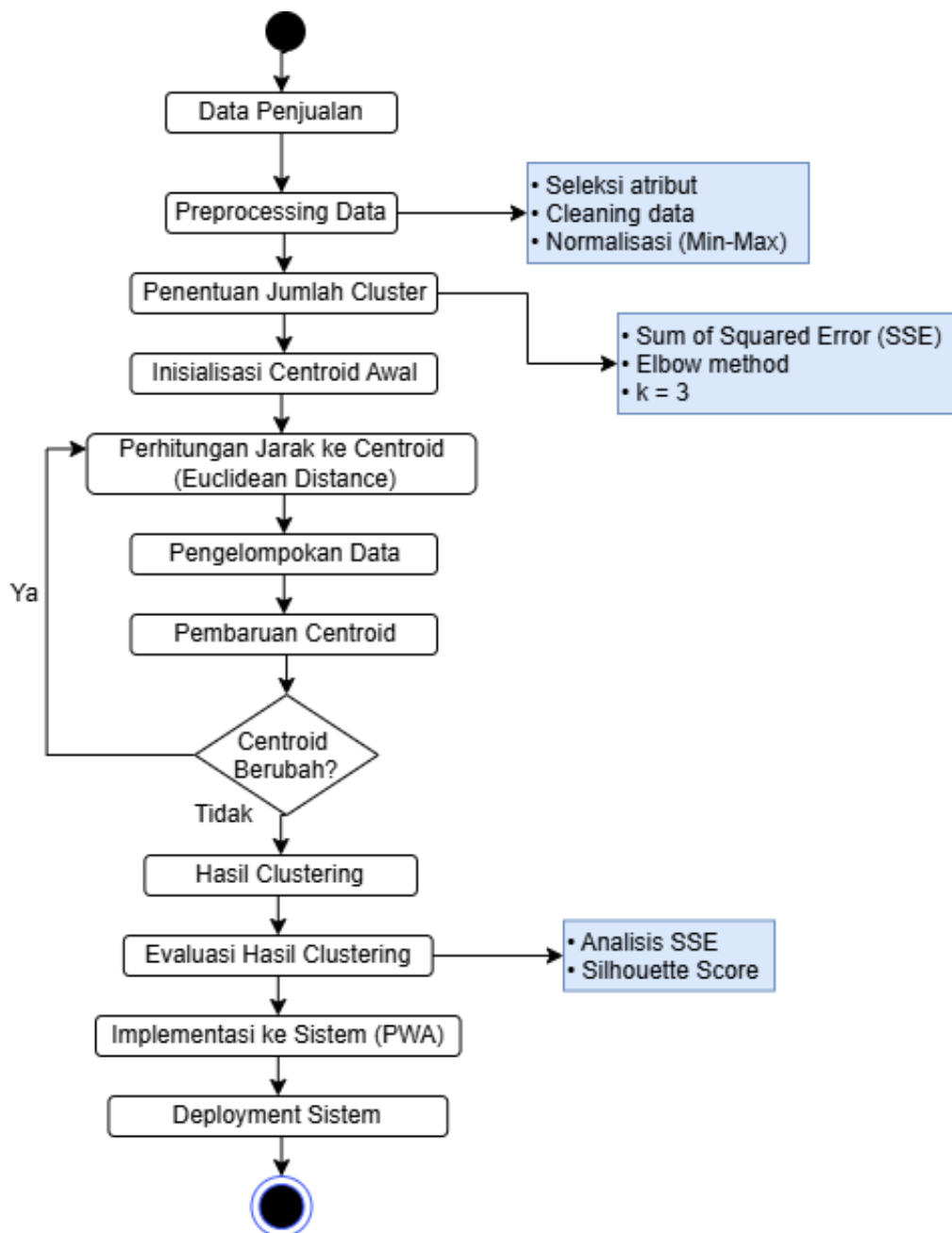
Hak Cipta Dilindungi Undang-Undang

1. Dilarang memperbanyak atau mendistribusikan dokumen ini untuk tujuan komersial tanpa izin tertulis dari penulis atau pihak berwenang.
 2. Penggunaan tanpa izin untuk kepentingan akademik, penelitian, dan pendidikan diperbolehkan dengan mencantumkan sumber.
 3. Plagiarisme juga dilarang dan dapat dikenakan sanksi.
 3. Universitas hanya berhak menyimpan dan mendistribusikan dokumen ini di repositori akademik, tanpa mengalihkan hak cipta penulis, sesuai dengan peraturan yang berlaku di Indonesia.

Universitas Islam Indragiri

Penentuan jumlah *cluster* ditetapkan sebanyak tiga ($k = 3$). Penetapan ini didasarkan pada kebutuhan Toko Azhar Tani Mandiri untuk mengelompokkan produk ke dalam tiga kategori berdasarkan jumlah penjualan (qty) dan total penjualan (total), yaitu produk paling laris, produk cukup laris, dan produk kurang laris. Kualitas hasil *clustering* dievaluasi menggunakan dua cara yaitu *Sum of Squared Error* (SSE) dan *Silhouette Score*. Nilai SSE digunakan untuk mengukur seberapa dekat jarak setiap data terhadap centroid *clusternya*, di mana semakin kecil nilai SSE maka semakin baik kualitas pengelompokan yang dihasilkan. *Silhouette Score* digunakan untuk mengukur seberapa baik pemisahan antar *cluster*, dengan nilai mendekati 1 menunjukkan pengelompokan yang sangat baik. Selain itu, sistem juga menampilkan grafik Elbow sebagai gambaran visual penurunan nilai SSE pada berbagai nilai k . Grafik Elbow bekerja dengan menghitung nilai SSE untuk beberapa kemungkinan jumlah *cluster*, misalnya $k = 2$, $k = 3$, $k = 4$, dan seterusnya, kemudian hasilnya digambarkan dalam bentuk grafik dengan sumbu horizontal menunjukkan jumlah *cluster* (k) dan sumbu vertikal menunjukkan nilai SSE. Titik optimal ditunjukkan pada bagian grafik yang membentuk pola menyerupai sudut siku, yaitu titik di mana penurunan nilai SSE mulai melambat secara signifikan. Grafik ini dapat digunakan sebagai informasi pendukung dalam memahami dan memverifikasi bahwa nilai $k = 3$ yang digunakan sudah menghasilkan pengelompokan yang baik.

3.1.4.2 Alur Kerja Algoritma K-Means Clustering



Gambar 3.2 Alur Kerja Algoritma K-Means Clustering

Gambar 3.2 menunjukkan alur kerja algoritma K-Means clustering yang diterapkan dalam penelitian ini, dimulai dari pengolahan data penjualan hingga implementasi hasil clustering ke dalam sistem berbasis PWA. Proses diawali dengan input data penjualan produk Toko Azhar Tani Mandiri tahun 2024 - 2025.





Data penjualan yang diperoleh kemudian melalui tahap preprocessing yang meliputi seleksi atribut, pembersihan data, dan normalisasi data menggunakan metode Min-Max Normalization. Setelah data siap digunakan, jumlah *cluster* ditentukan sebanyak tiga ($k = 3$) yang merepresentasikan produk paling laris, produk cukup laris, dan produk kurang laris berdasarkan jumlah penjualan (qty) dan total penjualan (total).

Selanjutnya, algoritma K-Means menginisialisasi centroid awal secara acak. Proses inti dimulai dengan menghitung jarak setiap data penjualan terhadap masing-masing centroid menggunakan metode *Euclidean Distance*. Kemudian hasil perhitungan jarak tersebut, setiap data akan dikelompokkan ke dalam *cluster* dengan jarak terdekat. Setelah proses pengelompokan dilakukan, centroid baru dihitung kembali berdasarkan nilai rata-rata data pada setiap *cluster*. Sistem kemudian melakukan pengecekan terhadap perubahan nilai centroid. Apabila centroid masih mengalami perubahan, maka proses iterasi dilanjutkan kembali ke tahap perhitungan jarak dan pengelompokan data. Proses ini dilakukan secara berulang hingga tidak terjadi perubahan signifikan pada nilai centroid. Ketika centroid sudah stabil, algoritma dinyatakan mencapai kondisi konvergensi dan menghasilkan hasil *clustering* akhir. Hasil *clustering* yang diperoleh selanjutnya dievaluasi menggunakan SSE dan *Silhouette Score* untuk memastikan kualitas pengelompokan data. Setelah proses evaluasi selesai, hasil *clustering* diimplementasikan ke dalam sistem berbasis PWA dan di-deploy agar dapat digunakan oleh pemilik Toko Azhar Tani Mandiri sebagai dasar keputusan pembelian stok produk.

1. Dilarang memperbanyak atau mendistribusikan dokumen ini untuk tujuan komersial tanpa izin tertulis dari penulis atau pihak berwenang.
2. Penggunaan tanpa izin untuk kepentingan akademik, penelitian, dan pendidikan diperbolehkan dengan mencantumkan sumber.
3. Plagiarisme juga dilarang dan dapat dikenakan sanksi.
3. Universitas hanya berhak menyimpan dan mendistribusikan dokumen ini di repositori akademik, tanpa mengalihkan hak cipta penulis, sesuai dengan peraturan yang berlaku di Indonesia.

3.1.5 Evaluasi Hasil Clustering

Evaluasi hasil *clustering* dilakukan untuk memastikan bahwa algoritma K-Means mampu mengelompokkan data penjualan produk secara tepat dan menghasilkan *cluster* yang berkualitas. Dalam penelitian ini, evaluasi hasil *clustering* dilakukan menggunakan dua pendekatan utama, yaitu *Sum of Squared Error* (SSE) dan *Silhouette Score*.

1. *Sum of Squared Error* (SSE) merupakan metrik yang digunakan untuk mengukur tingkat kekompakan *cluster* yang dihasilkan. SSE menghitung total kuadrat jarak antara setiap data dengan centroid *cluster* tempat data tersebut berada. Semakin kecil nilai SSE, maka semakin baik kualitas *clustering* karena data dalam satu *cluster* memiliki jarak yang lebih dekat dengan centroid-nya. Nilai SSE dihitung menggunakan rumus:

$$SSE = \sum_{i=1}^k \sum_{x \in C_i} d(x, C_i)^2 \quad (4)$$

Dimana:

k = data *cluster* ke k

C_i = *cluster* ke i

$x \in C_i$ = data ke- x dalam *cluster*

$d(x, C_i)$ = jarak data dengan centroid

2. *Silhouette Score* untuk mengukur kualitas pemisahan antar *cluster* dan tingkat kesamaan data dalam satu *cluster*. *Silhouette Score* menunjukkan seberapa baik suatu data ditempatkan pada *cluster* yang tepat dibandingkan dengan *cluster* lainnya. Nilai *Silhouette Score* berada pada rentang -1 hingga 1. Nilai yang mendekati 1 menunjukkan bahwa data telah dikelompokkan dengan sangat



Hak Cipta Dilindungi Undang-Undang

1. Dilarang memperbanyak atau mendistribusikan dokumen ini untuk tujuan komersial tanpa izin tertulis dari penulis atau pihak berwenang.
2. Penggunaan tanpa izin untuk kepentingan akademik, penelitian, dan pendidikan diperbolehkan dengan mencantumkan sumber.
3. Plagiarisme juga dilarang dan dapat dikenakan sanksi.

Universitas hanya berhak menyimpan dan mendistribusikan dokumen ini di repositori akademik, tanpa mengalihkan hak cipta penulis, sesuai dengan peraturan yang berlaku di Indonesia.



Hak Cipta Dilindungi Undang-Undang

1. Dilarang memperbanyak atau mendistribusikan dokumen ini untuk tujuan komersial tanpa izin tertulis dari penulis atau pihak berwenang.
 2. Penggunaan tanpa izin untuk kepentingan akademik, penelitian, dan pendidikan diperbolehkan dengan mencantumkan sumber.
 3. Universitas hanya berhak menyimpan dan mendistribusikan dokumen ini di repositori akademik, tanpa mengalihkan hak cipta penulis, sesuai dengan peraturan yang berlaku di Indonesia.

Universitas Islam Indragiri

baik, nilai mendekati 0 menunjukkan adanya tumpang tindih antar *cluster*, sedangkan nilai negatif menunjukkan kemungkinan kesalahan dalam pengelompokan data. Dengan demikian, hasil pengelompokan produk dapat digunakan sebagai dasar keputusan pembelian stok pada Toko Azhar Tani Mandiri.

3.1.6 Perancangan Sistem

Pada tahap ini, perancangan sistem divisualisasikan menggunakan diagram UML yang akan dibahas secara lebih rinci pada BAB IV. Diagram UML digunakan untuk menggambarkan fungsi-fungsi utama yang dapat dilakukan oleh pengguna sistem serta alur aktivitas sistem secara berurutan, dimulai dari login, proses upload data penjualan, preprocessing data, proses *clustering* menggunakan algoritma K-Means, penyajian hasil pengelompokan produk dalam bentuk visualisasi, hingga melihat laporan dan riwayat sesi *clustering* yang telah dilakukan sebelumnya.

3.1.7 Implementasi Sistem

Tahap implementasi merupakan tahap dari hasil analisis dan perancangan sistem yang telah dilakukan sebelumnya. Pada tahap ini, sistem dikembangkan menggunakan pendekatan *Waterfall* yang dilakukan secara bertahap dan terstruktur. Sebelum diimplementasikan ke dalam sistem, algoritma K-Means terlebih dahulu dihitung secara manual menggunakan Microsoft Excel sebagai langkah awal untuk memastikan kebenaran alur perhitungan mulai dari normalisasi Min-Max, perhitungan jarak *Euclidean*, pembaruan centroid, hingga pengelompokan data ke dalam tiga *cluster*. Setelah hasil perhitungan manual diperoleh, proses yang sama kemudian dijalankan menggunakan *Google Colab* berbasis Python untuk memvalidasi bahwa hasil dari kedua cara tersebut konsisten.



Hak Cipta Dilindungi Undang-Undang

1. Dilarang memperbanyak atau mendistribusikan dokumen ini untuk tujuan komersial tanpa izin tertulis dari penulis atau pihak berwenang.
Penggunaan untuk kepentingan akademik, penelitian, dan pendidikan diperbolehkan dengan mencantumkan sumber.

2. Penggunaan tanpa izin untuk kepentingan komersial atau pelanggaran hak cipta dapat dikenakan sanksi sesuai dengan UU Hak Cipta di Indonesia. Plagiarisme juga dilarang dan dapat dikenakan sanksi.

3. Universitas hanya berhak menyimpan dan mendistribusikan dokumen ini di repositori akademik, tanpa mengalihkan hak cipta penulis, sesuai dengan peraturan yang berlaku di Indonesia.

Barulah setelah hasil Excel dan *Google Colab* menunjukkan nilai yang sama, algoritma diimplementasikan ke dalam sistem menggunakan bahasa pemrograman PHP. Rumus matematika yang digunakan pada ketiga tahap adalah identik, perbedaannya hanya pada cara pengerjaannya yaitu manual di Excel, sintaks Python di *Google Colab*, dan sintaks PHP pada sistem yang dibangun, sehingga hasil yang dihasilkan sistem sama. Implementasi PWA mencakup *web app manifest* untuk mendukung fitur *installable app* sehingga sistem dapat diakses seperti aplikasi native pada perangkat pengguna, serta protokol HTTPS untuk keamanan. Sistem memerlukan koneksi internet untuk mengakses data penjualan dan menjalankan proses *clustering*.

3.1.8 Pengujian Sistem

Pengujian sistem mencakup uji coba langsung melalui aplikasi web untuk memastikan fungsi *clustering* berjalan baik dan hasil pengelompokan ditampilkan secara akurat dengan visualisasi yang interaktif. Pengujian fungsional dilakukan menggunakan metode *black-box testing* untuk memastikan seluruh fitur berfungsi dengan baik, meliputi fungsi upload data penjualan, proses *clustering*, visualisasi hasil (*scatter plot*, grafik Elbow, grafik SSE per iterasi), tampilan hasil pengelompokan produk, laporan, riwayat sesi, serta fitur PWA (*install to home screen*).