

## BAB II KERANGKA TEORITIS

### A. Landasan Teori

#### 1. Algoritma *K-Means*

Pada penelitian ini, dibutuhkan algoritma *K-Means* untuk pemetaan potensi perluasan pasar dalam pemberian pelatihan *digital marketing* pada usaha mikro, kecil, dan menengah. Menurut (Santoso, Azis, & Zohrahayaty, 2020, p. 122) *K-Means* merupakan salah satu algoritma *machine learning* untuk memecahkan masalah pengelompokan. Metode *K-Means* merupakan salah satu metode *clustering non-hirari* dimana algoritma *K-Means* berusaha mencari pusat dari kelompok datanya sebanyak iterasi perbaikan yang dilakukan. Metode *K-Means* dapat mengelompokkan suatu dataset ke dalam beberapa buah *cluster* (Primarta, 2020, p. 524)

Menurut (Arhami & Nasir, 2020, p. 148), langkah – langkah dasar algoritma *K-Means* adalah;

- (a) tentukan nilai *k cluster* sesuai dengan yang diinginkan;
- (b) pilih titik – titik atau sampel yang menjadi anggota *cluster* secara acak;
- (c) tentukan nilai *centroid* atau titik tengah dari *cluster* tersebut dengan rumus sebagai berikut;

$$M_k = \left(\frac{1}{n_k}\right) \sum_{i=1}^{Nk} x_{ik}$$

- (d) hitung *square error* untuk tiap *cluster*  $c_k$  yang merupakan jumlah kuadrat dari jarak *Euclidean* antara tiap sampel dalam  $c_k$  dan titik tengahnya (*centroid*); eror ini juga dikenal dengan nama *within cluster variation* (WCV), yaitu;

$$E_k^2 = \sum_{i=1}^{Nk} (X_{ik} - M_k)^2$$

- (e) selanjutnya jumlah dari keseluruhan eror dari *k-cluster* dihitung dengan rumus;

$$E_k^2 = \sum_{k=1}^k e_k^2$$

- (f) kelompokkan kembali semua sampel berdasarkan jarak minimum dari masing – masing pusat sehingga diperoleh distribusi baru dari sampel sesuai klasternya, untuk memperoleh distribusi sampel baru tersebut

dapat dilakukan dengan menghitung jarak dari masing – masing titik pusat dengan keseluruhan sampel; perhitungan jarak dapat menggunakan beberapa metode contohnya dengan *Euclidean Distance*;

$$d(p,q) = \sqrt{((p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_n - q_n)^2)}$$

- (g) tuliskan hasil anggota kluster baru sesuai dengan hasil yang diperoleh pada langkah ke-5; kemudian ulangi langkah ke-3 sampai beberapa iterasi sehingga nantinya ditemukan nilai total *square error* turun secara signifikan.

Menurut (Arhami & Nasir, 2020, p. 148), perhitungan *K-Means* dapat dijabarkan dengan contoh sebagai berikut ini;

$x_k$	p	q
$x_1$	0	2
$x_2$	0	0
$x_3$	1.5	0
$x_4$	5	0
$x_5$	5	2

- (a) misal pada kasus seperti diatas, nilai k yang diambil adalah  $K = 2$ , maka *cluster* yang diambil secara acak adalah sebagai berikut;

$$C_1 = \{ x_1, x_2, x_3 \}$$

$$C_2 = \{ x_3, x_5 \}$$

- (b) maka dapat ditentukan titik tengah dari kedua *cluster* diatas, yaitu;

$$M_1 = \left\{ \frac{0 + 0 + 5}{3}, \frac{2 + 0 + 0}{3} \right\} = \{1.66, 0.66\}$$

$$M_2 = \left\{ \frac{1.5 + 5}{2}, \frac{0 + 2}{2} \right\} = \{3.25, 1.00\}$$

- (c) *square error* diperoleh adalah;

$$e_1^2 = [(0 - 1.66)^2 + (2 - 0.66)^2 + [(0 - 1.66)^2 + (0 - 0.66)^2] + [(5 - 1.66)^2 + (0 - 0.66)^2]] = 19.36$$

$$e_2^2 = [(1.5 - 3.25)^2 + (0 - 1)^2 + [(5 - 3.25)^2 + (2 - 1)^2]] = 8.12$$

Sehingga *total square error*-nya yaitu;

$$E^2 = e_1^2 + e_2^2 = 19.36 + 8.12 = 27.48$$

(d) berikutnya adalah dilakukan redistribusi sampai kedalam kelas yang sesuai dengan hasil perhitungan jarak terpendek yang dimiliki oleh setiap sampel dengan jarak minimum dari titik pusat *cluster* yang ada akan menjadi *cluster* baru bagi sampelnya;

$$d(M_1, x_1) = \sqrt{(1.66 - 0)^2 + (0.66 - 2)^2} = 2.14$$

$$d(M_2, x_1) = \sqrt{(3.25 - 0)^2 + (1 - 2)^2} = 3.40$$

$$d(M_1, x_2) = \sqrt{(1.66 - 0)^2 + (0.66 - 0)^2} = 1.79$$

$$d(M_2, x_2) = \sqrt{(3.25 - 0)^2 + (1 - 0)^2} = 3.40$$

$$d(M_1, x_3) = \sqrt{(1.66 - 1.5)^2 + (0.66 - 0)^2} = 0.83$$

$$d(M_2, x_3) = \sqrt{(3.25 - 1.5)^2 + (1 - 0)^2} = 2.01$$

$$d(M_1, x_4) = \sqrt{(1.66 - 5)^2 + (0.66 - 0)^2} = 3.41$$

$$d(M_2, x_4) = \sqrt{(3.25 - 5)^2 + (1 - 0)^2} = 2.01$$

$$d(M_1, x_5) = \sqrt{(1.66 - 5)^2 + (0.66 - 2)^2} = 3.60$$

$$d(M_1, x_1) = \sqrt{(3.25 - 5)^2 + (1 - 2)^2} = 2.01$$

untuk memudahkan dalam penentuan klasternya maka hasil perhitungan diatas dapat dibuat dalam bentuk tabel seperti berikut, dan jarak minimum dari titik tengah klaster ke sampel akan menjadi kelas sampel;

$x_k$	$d(M_1, x_{k2})$	$d(M_2, x_{k2})$	<b>Cluster</b>
$x_1$	2.14	3.40	$c_1$
$x_2$	1.79	3.40	$c$
$x_3$	0.83	2.01	$c$
$x_4$	3.41	2.01	$c_4$
$x_5$	<b>3.60</b>	<b>2.01</b>	$c$

kedua klaster tersebut dihitung kembali titik tengah dan total dari *square error*-nya untuk memastikan apakah perlu dilanjutkan ke iterasi berikutnya atau tidak; titik tengah dari kedua klaster tersebut adalah sebagai berikut;

$$M_1 = \left\{ \frac{0 + 0 + 1.5}{3} \right\}, \left\{ \frac{2 + 0 + 0}{3} \right\} = \{0.5, 0.66\}$$

$$M_2 = \left\{ \frac{5 + 5}{2} \right\}, \left\{ \frac{0 + 2}{2} \right\} = \{5, 1.00\}$$

dan *square error*-nya adalah  $e_1^2 = 4.17$  dan  $e_2^2 = 2$  sehingga totalnya diperoleh;

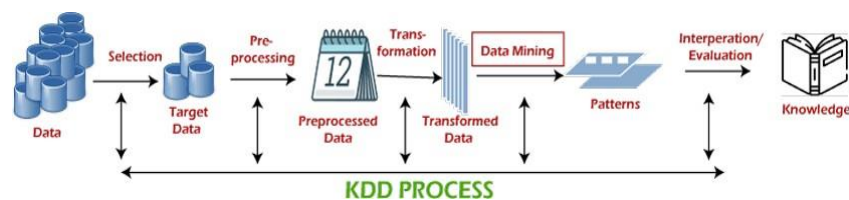
$$E^2 = e_1^2 + e_2^2 = 4.17 + 2 = 6.17$$

jika dilihat dari total *square error*-nya yang turun begitu signifikan dari 27.48 ke 6.17, maka untuk kasus sederhana ini iterasi pertama sudah

cukup untuk mengambil kesimpulan bahwa klaster yang dihasilkan adalah  $C_1 = \{x_1, x_2, x_3\}, C_2 = \{x_4, x_5\}$ .

## 2. Pemahaman Data Mining

*Data Mining* adalah proses penggalian informasi dan pola yang bermanfaat dari suatu data yang sangat amat besar. lainnya (Amna, 2023) Proses data mining terdiri dari berbagai pengumpulan data, ekstraksi data, analisa data, dan statistik data. Menurut (Fayyad, 1996, p.6) istilah "data mining" dan "knowledge discovery in databases (KDD)" digunakan secara bersamaan untuk menjelaskan proses penggalian informasi tersembunyi dalam suatu basis data yang sangat besar, adapun pada Gambar 2.1 sebagai berikut;



Gambar 2.1 Bagan Data Mining

Adapun Langkah-langkah knowledge discovery in databases (KDD) yang akan dijelaskan sebagai berikut:

- (1) *Data Selection*, Sebelum dimulai tahap penggalian informasi dalam KDD dari data seleksi, yang harus dilakukan dari data yang dipilih sebagai hasil dari proses penggalian data disimpan dalam suatu berkas yang berbeda dari basis data operasional;
- (2) *Pre-processing*, proses cleaning mencakup menghapus duplikat data, memeriksa data yang tidak konsisten, dan memperbaiki kesalahan data seperti kesalahan cetak (tipografi), dan proses pengembangan dilakukan untuk memperkaya data yang sudah ada dengan data atau informasi lain yang relevan dan diperlukan KDD, seperti data atau informasi dari sumber luar;
- (3) *Coding transformasi*, proses mengubah data yang telah dipilih sehingga dapat digunakan untuk proses data mining;
- (4) *Data mining*, proses mencari pola atau informasi yang menarik dalam data terpilih dengan teknik, metode, atau algoritma yang digunakan dalam data mining sangat beragam, dan pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan;

(5) Interpretation/Evaluation, tahap yang memeriksa apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya.

### 3. Sistem Pendukung Keputusan (SPK)

Menurut Wijoyo, H. dalam (Yuswardi, dkk., 2022, p. 1) menyatakan bahwa pengambilan keputusan dilakukan dengan beberapa perhitungan dimana termasuk mengumpulkan informasi, menentukan tujuan, mengidentifikasi alternatif, menetapkan kriteria dan memilih pilihan terbaik yang tersedia. Menurut (Yuswardi, dkk., 2022, p. 1) mengutarakan bahwa sistem pendukung keputusan merupakan suatu proses berpikir dan memutuskan berbagai pilihan dari pemilihan alternatif untuk meramalkan keadaan masa depan yang dibutuhkannya sebuah pemahaman teori atau metode mengenai pendukung keputusan. Metode K-Means digunakan pada penelitian ini sebagai sistem pendukung keputusan.

Sistem pendukung keputusan dapat meningkatkan efektivitas dan efisiensi para pengambil keputusan dalam membantu pengambilan keputusan (Wiji Setyaningsih, 2015, p. 12). Dalam pengambilan keputusan terdapat tahapan yang harus dilalui agar dapat menghasilkan sebuah keputusan terbaik. Adapun tahapan dalam pengambilan keputusan menurut (Jeperson Hutahaeen, dkk., 2023, p. 4-5):

- (a) *Intelligence phase*, proses dalam menelusuri dan mengenali permasalahan;
- (b) *design phase*, proses pengembangan dalam mencari solusi atau alternatif yang mungkin dapat diambil dengan adanya verifikasi dan validasi untuk mengetahui tingkat akurasi model yang digunakan;
- (c) *choice phase*, dilakukan sebuah pemilihan dari solusi alternatif yang telah ada dengan memperhatikan kriteria yang tepat dengan tujuan utamanya;
- (d) *implementation phase*, penerapan dilakukan sesuai dengan rancangan yang telah dibuat sebelumnya.

### 4. Unified Modeling Language (UML)

Dalam perancangan sistem, dibutuhkan suatu pemodelan yang dapat menggambarkan sistem. (Fitria Nur Hasanah dan Rahmania Sri Untari, 2020, p. 64) menjelaskan bahwa *Unified Modelling Language (UML)* adalah bahasa yang digunakan untuk mendefinisikan kemampuan sebuah sistem, desain, analisis serta gambaran mengenai *arsitektur* pada sebuah sistem. Menurut (A. Suhendar dan Hariman Gunadi, 2002, p. 34) *Unified Modelling Language (UML)* merupakan bahasa pemodelan dari sebuah sistem untuk

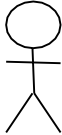
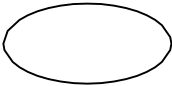

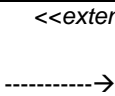
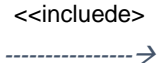
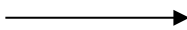
menggambarkan, membangun dan mendokumentasikan sistem yang besar dan rumit yang terdapat notasi untuk membantu dalam menjelaskan secara *visual* mengenai komponen yang terdapat pada pemodelan.

Menurut (Kusnardi, supiadi, Syabaniah, & Oktapiani, 2020), uml memiliki beragam diagram yang dapat mewakili berbagai sudut pandang dari sistem/ perangkat lunak yang akan dibangun, yaitu;

**(a) use case diagram**

menjelaskan sebuah interaksi antara satu atau lebih *actor* dengan sistem informasi yang akan dibuat serta dapat mendeskripsikan tipe interaksi antara si pengguna sistem dengan sistemnya; , berikut merupakan simbol pada use case diagram;


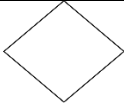
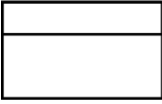
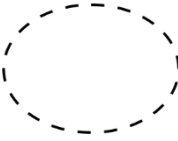
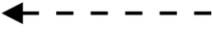
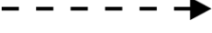
Tabel 2.1 Use case diagram


Simbol	Nama	Keterangan
	ACTOR	Actor yang menggambarkan seseorang berinteraksi dengan sistem
	USECASE	usecase menggambarkan apa yang dikerjakan sistem, biasanya menggunakan kata kerja.
	ASOSIASI/ ASSOCIATION	menghubungkan antara aktor dengan usecase
	EKSTENSI/ EXTEND	Digunakan sebagai perluasan dari usecase lain jika kondisi terpenuhi
	INCLUDE	menghungkan usecase dengan usecase yang prosesnya diaruskan atau menjadi syarat yang harus dipenuhi
	GENERALISASI/ GENERALIZATION	Actor yang berhubungan secara pasif dengan sistem

**(b) class diagram**

*class diagram* menjelaskan mengenai aliran data pada sebuah program yang dimana menggambarkan struktur sistem dengan mendefinisikan *class object* yang menyusun sistem dan hubungan yang terdapat pada *class object* tersebut, berikut merupakan simbol pada *class diagram*;

Tabel 2.2 Class Diagram



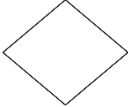
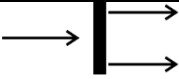


Simbol	Nama	Keterangan
	<i>Generalization</i>	Hubungan dimana objek anak (descendent) berbagi perilaku dan struktur data dari objek yang ada di atasnya objek induk (ancestor)
	<i>Navy Association</i>	Upaya untuk menghindari asosiasi dengan lebih dari 2 objek
	Class	Himpunan dari objek-objek yang berbagi atribut serta operasi yang sama
	<i>Collaboration</i>	Deskripsi dari urutan aksi-aksi yang ditampilkan sistem yang menghasilkan suatu hasil yang terukur bagi suatu actor.
	<i>Realization</i>	Operasi yang benar-benar dilakukan oleh suatu objek
	<i>Dependency</i>	Hubungan dimana perubahan yang terjadi pada suatu elemen mandiri (independent) akan mempengaruhi elemen yang bergantung

Simbol	Nama	Keterangan
	<i>Association</i>	Yang menghubungkan antara objek satu dengan objek lainnya.

**(c) activity diagram**

*activity* diagram didefinisikan sebagai alur dari aktivitas yang terjadi pada sebuah sistem yang menggambarkan kegiatan yang akan dilakukan pada sistem atau program tersebut, berikut merupakan simbol pada *activity* diagram;

Tabel 2.3 Activity Diagram



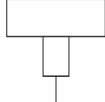
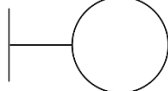
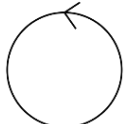
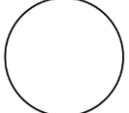


Simbol	Nama	Keterangan
	Status Awal/Initial	Status awal aktivitas sistem
	Activity	Aktivitas yang dilakukan sistem, aktivitas biasanya diawali dengan kata kerja.
	<i>Decision</i>	Asosiasi percabangan dimana lebih dari satu aktivitas digabungkan menjadi satu.
	Join	Asosiasi penggabungan dimana lebih dari satu aktivitas lebih dari satu.
	Status Akhir/Final	Status akhir yang dilakukan sistem,.
	Swimline	Memisahkan organisasi bisnis yang bertanggung jawab terhadap aktivitas yang terjadi

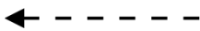
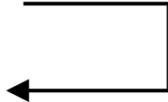
**(d) sequence diagram**

*sequence* diagram memberi gambaran bagaimana interaksi antar *object* yang terlibat yang diilustrasikan dengan pesan terkirim maupun diterima dalam sekuensinya, adapun simbol pada *sequence* diagram sebagai berikut;



Tabel 2.4 Sequence Diagram

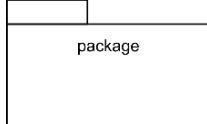
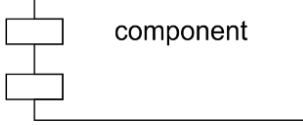

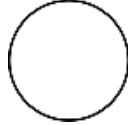
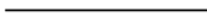
Simbol	Nama	Keterangan
	<i>Actor</i>	Mempresentasikan entitas yang berada diluar sistem dan berinteraksi diluar sistem.
	<i>Lifeline</i>	Menghubungkan objek selama sequence (message dikirim atau diterima).
	<i>General</i>	Merepresentasikan entitas tunggal dalam sequence
	<i>Boundary</i>	Berupa tepi dari sistem, seperti user interface dan alat yang berinteraksi dengan yang lain.
	<i>Control</i>	Elemen mengatur aliran dari informasi untuk sebuah scenario
	<i>Entity</i>	Elemen yang bertanggung jawab menyimpan informasi
	<i>Activation</i>	Suatu titik dimana sebuah objek mulai berpartisipasi dalam sebuah sequence yang menunjukkan sebuah objek atau mengirim atau menerima objek.
	<i>Message Entry</i>	Berfungsi untuk menggambarkan pesan/ hubungan antar objek yang menunjukkan urutan kejadian yang terjadi.

Simbol	Nama	Keterangan
	<i>Message to Self</i>	Simbol ini menggambarkan pesan/hubungan objek itu sendiri, yang menunjukkan urutan kejadian yang terjadi.
	<i>Message Return</i>	Menggambarkan hasil dari pengiriman message yang digambarkan dengan arah dari kanan ke kiri.

### (e) Component Diagram

Diagram component atau komponen diagram ialah sesuatu yang digunakan untuk memperlihatkan sebuah organisasi serta keterkaitan dan ketergantungan diantara kumpulan komponen dalam sebuah sistem, adapun simbol pada komponen diagram sebagai berikut;

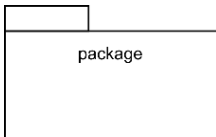
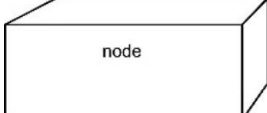
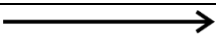

Tabel 2.5 Component Diagram

Simbol	Nama	Keterangan
	<i>Package</i>	<i>Package</i> merupakan sebuah bungkusan dari satu atau lebih komponen
	<i>Component</i>	Komponen sistem
	<i>Dependency</i>	Kebergantungan antar komponen, arah panah mengarah pada komponen yang dipakai.
	<i>Interface</i>	Sebagai antarmuka komponen agar tidak mengakses langsung komponen.
	<i>Link</i>	Relasi antar komponen

(f) **deployment diagram**

*deployment diagram* digunakan untuk menggambarkan mengenai *software* maupun *hardware* pada sistem yang menunjukkan konfigurasi komponen pada proses pemakaian aplikasi tersebut, adapun simbol yang terdapat pada *deployment diagram* sebagai berikut;

Tabel 2.6 *Deployment Diagram*

Simbol	Nama	Keterangan
	<i>Package</i>	Package merupakan sebuah bungkusan dari satu atau lebih node
	<i>Node</i>	Node biasanya mengacu pada perangkat keras ( <i>hardware</i> ), perangkat lunak yang tidak dibuat sendiri ( <i>software</i> ), jika di dalam <i>node</i> disertakan komponen untuk mengkonsistenkan rancangan maka komponen yang diikutsertakan harus sesuai dengan komponen yang telah didefinisikan sebelumnya pada diagram komponen.
	<i>Dependency</i>	Kebergantungan antar <i>node</i> , arah panah mengarah pada <i>node</i> yang dipakai.
	<i>Link</i>	Relasi antar <i>node</i> .

5. **Pengembangan System Development Life Cycle (SDLC)**

Penelitian Dalam hal ini penelitian yang dilakukan menggunakan sebuah model pengembangan agar proses yang terjadi dalam penelitian ini lebih terstruktur. Menurut (Rosa A.S dan M. Shalahuddin, 2013, p. 26) System Development Life Cycle (SDLC) merupakan suatu pengembangan pada sistem yang sudah ada sebelumnya dengan memiliki tahapan-tahapan yang

harus dilalui agar terciptanya sebuah sistem yang berkualitas. Pada System Development Life Cycle (SDLC) terdapat beberapa metode, salah satu metode yang digunakan pada penelitian ini yaitu metode *prototyping*. Menurut (Abdul Kadir, 2014, p. 29) *prototyping* merupakan metode dalam pengembangan sebuah sistem yang dilakukannya pendekatan untuk menghasilkan spesifikasi yang terperinci sebelum pengguna melakukan evaluasi pada sistem tersebut.

Menurut (Hartono,2020, p. 73) tahapan – tahapan SDLC sebagai berikut;



Gambar 2.2 Metode SDLC

adapun tahapan-tahapan pada gambar (2.2) dijelaskan sebagai berikut:

- (1) fase Perencanaan Sistem, pada tahapan ini yaitu membentuk suatu struktur kerja strategis secara luas dan pandangan sistem informasi baru yang jelas yang kemudian akan memenuhi kebutuhan-kebutuhan pemakainya. Sistem kemudian dievaluasi dan dipisahkan berdasarkan prioritas berdasarkan indikator kebutuhan. Yang memiliki prioritas tertinggi selanjutnya akan dipilih untuk pengembangan, penyediaan sumber daya baru, dan penyediaan sumberdaya untuk keperluan pengembangan sistem. Rencana kerja yang matang juga akan disusun untuk menjalankan tahapan-tahapan lainnya; Kemudian hasil dari tahapan ini adalah, langkah-langkah detail terkait rencana kerja dan penugasan untuk anggota tim yang bersangkutan
- (2) fase Analisis Sistem, pada tahap ini kemudian dilakukan proses penilaian, identifikasi dan evaluasi terkait komponen dan hubungan

timbal-balik yang berkaitan dengan pengembangan sistem seperti, definisi masalah, tujuan, kebutuhan, prioritas dan kendala-kendala system, ditambah identifikasi anggaran, keuntungan, serta estimasi jadwal untuk solusi yang dinilai berpotensi. Fase analisis sistem disebut sebagai fase profesional sistem melakukan kegiatan analisis sistem;

(3) fase Perancangan Sistem Awal, pada tahapan ini kemudian dibentuk berbagai alternatif perancangan konseptual dari sudut pandang pemakai. Alternatif ini merupakan perluasan dari kebutuhan pemakai. Alternatif perancangan konseptual memungkinkan manajer atau pemakai untuk memilihrancangan terbaik yang cocok untuk kebutuhan mereka. Pada fase ini selanjutnya analisis sistem dimulai dari merancang proses dengan mengidentifikasi berbagai laporan dan keluaran yang akan dihasilkan oleh sistem yang diusulkan;

#### **6. Prototyping dan Clustering**

Menurut Roger S. Presman (2010, p.43 – 44) menyatakan bahwa *prototyping* dimulai dengan *communication* yaitu komunikasi bertemu dengan pihak – pihak terkait dengan mendefinisikan tujuan keseluruhan dari perangkat lunak ini melibatkan diskusi dan pembicaraan untuk memahami apa yang ingin dicapai dengan pengembangan perangkat lunak tersebut; *Quick Plan* yaitu perencanaan secara cepat setelah komunikasi; selanjutnya *Modeling Quick Design* yaitu membuat model dan desain cepat ini fokus pada representasi aspek – aspek perangkat lunak yang akan terlihat atau dirasakan oleh pengguna akhir contohnya adalah tata letak antar muka manusia (human interface layout) atau tampilan output; *Construction of prototype* hasil dari *quick design* adalah pembuatan *prototype*; *prototype* adalah versi awal dari perangkat lunak yang dirancang untuk memberikan gambaran nyata tentang bagaimana perangkat lunak tersebut berfungsi. Berikut ini model *prototyping* menurut (Rosa A.S dan M. Shalahuddin, 2013, p. 32) yang terdiri dari 3 tahapan ditunjukkan pada gambar 2.3, sebagai berikut;

Gambar 2.3 Model Prototyping



Sumber : (Rosa A.S dan M. Shalahuddin, 2013, p. 32)

- (1) melakukan pengumpulan kebutuhan dari sebuah sistem dengan cara mendengarkan pelanggan untuk mengetahui kebutuhan bagaimana sistem yang sedang berjalan dengan melakukan analisis kebutuhan dari pengguna;
- (2) dilakukannya perancangan dan pembuatan *prototyping* sesuai dengan kebutuhan pengguna yang telah dijabarkan sebelumnya;
- (3) dilakukannya pengujian *prototyping* oleh pengguna kemudian pengguna memberikan evaluasi kekurangan dan kebutuhan apa saja yang dibutuhkan oleh pengguna sehingga pada saat pengujian masih terdapat kekurangan dan kebutuhan maka kembali kepada tahap mendengarkan pelanggan untuk memperbaiki *prototyping* yang sudah tersedia; *prototyping* sebuah sistem yang dapat menyediakan dan mendemonstrasikan fungsi-fungsi yang terdapat serta mungkin dapat dilakukan pengujian desain sistem tersebut dapat dikatakan sebagai mock-up.

Untuk melakukan pengelompokkan, dibutuhkan suatu teknik yang tepat sesuai dengan hasil yang diharapkan. Menurut (Jollyta, Ramdhan, Zarlis, 2020, p. 53), *clustering* adalah suatu model yang dilakukan dengan cara melakukan segmentasi terhadap populasi yang heterogen ke dalam sejumlah *cluster* yang homogeny; *clustering* dilakukan dengan cara mengumpulkan data yang serumpun dari sebuah data set yang lebih besar; dengan *clustering*, kelompok minoritas yang memiliki kemiripan entitas; *clustering* dapat digunakan untuk mendeteksi secara otomatis *cluster* dari record yang berdekatan yang memiliki pengertian tertentu didalam keseluruhan variabel;

terdapat 4 pendekatan untuk permasalahan *clustering*, yaitu, *partition clustering*, *grid – base clustering*, *hierarchial clustering*, dan *density – based clustering*. Pada penelitian ini, model *clustering* digunakan untuk mengelompokkan data usaha mikro, kecil, dan menengah dengan pendekatan *partition clustering*.

## 7. Pemograman

Bahasa pemrograman yang tepat dibutuhkan dalam membangun sebuah aplikasi yang terdapat pada program komputer guna menghasilkan sebuah aplikasi yang sesuai dengan yang diharapkan, bahasa pemrograman yang digunakan dalam penelitian ini yaitu *python*. Menurut (Adbul Kadir, 2014, p. 192) bahasa pemrograman merupakan sekumpulan bahasa perintah yang dimengerti oleh komputer dalam hal berkomunikasi untuk menjalankan perintah tertentu yang pada akhirnya perintah tersebut dapat dijalankan oleh komputer. (Budiharto, 2018, p. 13) menjelaskan bahwa *Python* merupakan tools perangkat lunak yang dapat dilakukan untuk perhitungan numerik, visualisasi dan menganalisis dengan memiliki berbagai library pendukung. Bahasa pemrograman *Python* menunjang untuk pemrograman berorientasi objek dan pemrograma fungsional.

Streamlit merupakan sebuah *framework open-source* untuk *machine learning* dan *data science*. (Menurut putranto, dkk, 2023) memberikan pemahaman tentang streamlit yang merupakan suatu pustaka Python yang memungkinkan pengguna mengubah skrip data menjadi aplikasi berbasis web interaktif; Selain open source, streamlit juga mempromosikan open sharing, sehingga memudahkan pendistribusian dan berbagi aplikasi yang dibuat menggunakan library. streamlit merupakan kerangka kerja yang memungkinkan pengguna untuk membuat aplikasi web interaktif berbasis data menggunakan dengan menggunakan python (Richards, 2023, p. 23).

## 8. Pengujian Black Box

Terdapat teknik yang dibutuhkan dalam rancangan suatu program, teknik tersebut adalah *black-box testing*. Pengujian *black box* dilakukan untuk mengetahui fungsi-fungsi yang terdapat pada program sesuai dengan spesifikasi yang dibutuhkan. *Black box* testing merupakan pengujian yang menganalisa hasil pemeriksaan melewati uji coba data dan mengetahui kegunaan yang terdapat pada perangkat lunak (Emi Sita Eriana, 2022, p. 172); *Black box* testing dikerjakan dengan hanya menganalisa hasil pengecekan melewati uji coba dan menganalisa kegunaan perangkat lunak.

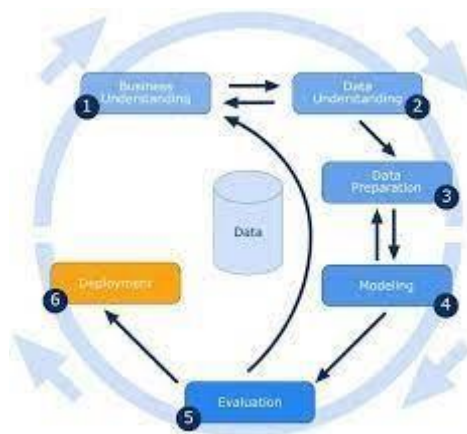
## 9. Post – Study System Usability Questionnaires (PSSUQ)

Pengujian dengan *post–study system usability* Teknik untuk melakukan pengujian dengan melihat penilaian kepuasan oleh pengguna dan untuk mengetahui adanya signifikansi perbedaan tingkat kesulitan pengguna terhadap sistem ataupun aplikasi. Menurut (Lewis, 2012, p. 1267) *Post-Study System Usability Questionnaire (PSSUQ)* merupakan sebuah kuesioner yang dirancang untuk menilai kepuasan yang dirasakan pengguna terhadap sistem atau aplikasi computer; Butir pertanyaan *PSSUQ* menghasilkan empat nilai, satu keseluruhan dan tiga subskala, nilai-nilai tersebut adalah:

- (a) *Overall*, keseluruhan: rata-rata tanggapan untuk pertanyaan 1 hingga 16;
- (b) *System Quality (SysQual)*, kualitas sistem: rata-rata dari pertanyaan 1 hingga 6;
- (c) *Information Quality (InfoQual)*, kualitas informasi: rata-rata dari pertanyaan 7 hingga 12;
- (d) *Interface Quality (IntQual)*, kualitas antarmuka: rata-rata dari pertanyaan 13 hingga 15.

## 10. Tahapan Crisp-DM

Crisp-DM salah satu kerangka proses dalam data mining. Cross-Industry Standard Process for Data Mining (CRISP-DM) merupakan strategi umum untuk memecahkan masalah bagi bisnis atau unit penelitian (Kusrini & Lutfi, 2009, p. 8). Menurut (Larose, 2006, p. 9) siklus hidup dari CRISP-SM terdiri dari 6 fase, sebagai berikut;



Gambar 2.4 Alur Crisp-DM

1. Business Understanding phase (fase pemahaman bisnis) adalah memahami objek bisnis, bagaimana membangun atau mendapatkan



- data, dan bagaimana mencocokkan tujuan pemodelan dengan tujuan bisnis sehingga model terbaik dapat dibuat,
2. Data Understanding Phase (Fase Pemahaman Data) adalah memberikan fondasi analitik untuk penelitian dengan membuat ringkasan dan menemukan masalah potensial dalam data,
  3. Data Preparation Phase (Fase Persiapan Data) adalah memastikan bahwa data tepat untuk algoritma yang digunakan, tahapan ini sering ditinjau kembali saat terjadi masalah selama pembangunan model, sehingga dilakukan iterasi sampai data cocok,
  4. Modelling Phase (Fase Pemodelan) merupakan tahapan yang dilakukan menggunakan teknik statistika dan Machine Learning untuk menentukan teknik data mining, alat bantu, dan algoritma data mining yang akan digunakan.
  5. Evaluation (Pengujian) adalah melakukan interpretasi terhadap hasil data mining yang dihasilkan dalam proses pemodelan pada tahap sebelumnya, Ini dilakukan dengan mengevaluasi model yang digunakan pada tahap sebelumnya agar model yang dipilih sesuai dengan tujuan yang ingin dicapai pada tahap pertama,
  6. Deployment (Penyebaran) merupakan menggabungkan metode untuk menghasilkan nilai model, mengubah skor keputusan, dan mengintegrasikan keputusan ke dalam sistem operasional.

## B. Tinjauan Pustaka

Dalam pelaksanaan penelitian, dilakukan proses penggalian informasi berdasarkan penelitian yang telah dilakukan sebelumnya. Penelitian mengenai pengelompokan UMKM sudah pernah dilakukan sebelumnya, begitu juga dengan penelitian yang menggunakan algoritma *K-Means*. Penelitian tersebut berkontribusi sebagai bahan perbandingan dan sumber pengetahuan penelitian yang dilaksanakan. Berikut adalah penelitian yang berkontribusi pada penulisan penelitian ini;

- (a) **Penentuan Jumlah Klaster Terbaik Pada K-Means Dalam Melihat Pola Klustering Data Mahasiswa Yang Telah Lulus** (Aisya Basri, Abdul Mubarak, Hairil Kurniadi Siradjuddin, Saiful Do. Abdullah, 2023), penelitian ini menggunakan Metode K-Means hasil analisis dan penerapan metode menunjukkan bahwa menggunakan metode Elbow dapat digunakan untuk menentukan jumlah cluster terbaik pada metode K-Means dengan studi kasus clustering data mahasiswa yang telah lulus, menghasilkan 2 cluster terbentuk

berdasarkan penerapan metode pada data dengan menggunakan dua atribut yaitu, nilai IPK dan jumlah SKS. Berdasarkan hasil perhitungan SSE maka didapatkan bahwa jumlah kluster terbaik pada penelitian ini ialah 2 kluster dengan nilai selisih SSE sebesar 4611.379920 dan berhasil membentuk garis siku pada grafik. Pengelompokan data berdasarkan jumlah kluster optimal dalam melakukan klustering waktu kelulusan terdiri dari kluster 1 sebanyak 199 data sebagai kluster tepat waktu dan kluster 2 sebanyak 41 data sebagai kluster tidak tepat waktu.

- (b) **Implementasi Metode *K-Means Clustering* dalam Analisis Persebaran UMKM di Jawa Barat** (Nursyifa, Resti Noor Fahmi, 2021), penelitian ini hasil evaluasi *clustering*nya dengan menerapkan algoritma *K-Means* yang dilakukan dengan membandingkan dua metode evaluasi menggunakan *silhouette coefficient* yang di mana didapatkan nilai sebesar 0,73 yang berarti masuk ke dalam kriteria kluster terbaik, Sedangkan dengan menggunakan metode *davies bouldin index* (DBI) didapatkan nilai sebesar 0,29 yang berarti masuk ke dalam kategori baik; dari hasil perbandingan kedua evaluasi *cluster* tersebut menunjukkan hasil *cluster* dengan menerapkan algoritma *K-Means* yang terbentuk baik;
- (c) **Strategi Pemasaran Produk Industri Kreatif Menggunakan Algoritma *K-Means Clustering* Berbasis Particle Swarm Optimization**(Oding Herdiana, Shanti Maulani, Eryan Ahmad Firdaus, 2021), penelitian mengelompokkan setiap atribut kontinu pada Data UMKM dan pemilihan jumlah *cluster* dapat mempengaruhi hasil akurasi dan waktu pemrosesan yang diperoleh, dalam 4 percobaan yaitu dengan menggunakan 2,3,4 dan *cluster*, jumlah *k-cluster* = 2 memberikan hasil akurasi dan waktu pemrosesan yang lebih baik dibanding jumlah *cluster* yang lain;
- (d) **Penerapan *Clustering K-Means* Untuk Menentukan Pengaruh Media Sosial Facebook Terhadap Usaha Mikro kecil dan Menengah (UMKM) Di Kecamatan Pekanbaru Kota**(Wisti Aristika, Wirajaya Hartono, 2020), pada penelitian ini terdapat kesamaan antara hasil manual dengan aplikasi rapidminer 5.3 yaitu *cluster* 0 sebanyak 22 item, dan *cluster* 1 sebanyak 38 item, metode *Clustering K-Means* dapat membantu mengelompokkan data untuk mengetahui pengaruh media social facebook terhadap peningkatan pendapatan UMKM khususnya di daerah Pekanbaru, penerapana algoritma *K-Means* membagi dataset menjadi 2 kelompok katagori yaitu kurang berpengaruh dan sangat berpengaruh dimana masing-masing *cluster* memiliki nilai rata-rata berbeda;

- (e) **Analisis Persebaran UMKM Kota Malang Menggunakan Cluster K-Means**(Puntoriza, Charista Fibriani, 2020), pada penelitian ini mengelompokkan UMKM di tiap Kecamatan di Kota Malang menjadi 3 *cluster* menggunakan metode *cluster K-Means*, hasil perhitungan menunjukkan kecamatan Blimbing masuk pada *cluster* 1 yaitu *cluster* tinggi, Kecamatan Klojen masuk pada *cluster* 2 yaitu *cluster* sedang, dan kecamatan Kedungkandang, Sukun, Lowokwaru pada *cluster* 3 yaitu *cluster* rendah; kecamatan tersebut akan mendapat perhatian pemerintah terkait penetapan strategi pemasaran untuk mengembangkan pasar berdasarkan tingkat prioritas masing-masing *cluster*;
- (f) **Penerapan Algoritma K-Means Clustering untuk Pengelompokan UMKM Menggunakan Rapidminer**(Wahyu Sudrajat, Idahm Cholid, Johannes Petrus, 2022), pada penelitian ini Tahapan penelitian yang dilakukan diantaranya adalah business understanding phase, data understanding phase, data processing phase, Modeling Phase, evaluation phase dan desimination phase. Pada pengujian 15 data usaha, penelitian ini berhasil menerapkan algoritma *K-Means Clustering* untuk mengidentifikasi dan mengklasifikasi UMKM dengan hasil pengujian sebesar 53% data ke *cluster* 1 sebanyak 8 data, 40% data ke *cluster* 2 sebanyak 6 data dan 7% data ke *cluster* 3 sebanyak 1 data, hasil perhitungan ini juga telah diuji coba menggunakan software rapidminer dan menghasilkan data yang sama;
- (g) **Analisis Klaster Kinerja Usaha Kecil dan Menengah Menggunakan Algoritma K-Means Clustering**(Dona Marcelina, Annisa Kurnia, Terttiaavini, 2023), pada penelitian ini menunjukkan bahwa pendekatan dengan menggunakan algoritma *K-Means* dapat membantu dalam mengelompokkan UKM berdasarkan karakteristik tertentu penerapan algoritma *K-Means Clustering* menggunakan aplikasi *KNIME* memberikan pendekatan terstruktur dalam mengelola data sehingga memungkinkan pengelompokan data kinerja UKM menjadi kelompok yang saling berhubungan dan identifikasi pola yang mungkin tidak terlihat sebelumnya, evaluasi model dengan menggunakan *node Regression Predictor* dan *node Numeric Score* memberikan gambaran tentang sejauh mana model klastering UKM dapat memprediksi secara akurat, Visualisasi data menggunakan *Scatter Plot* memungkinkan pemahaman yang lebih baik tentang pola klaster yang dihasilkan;
- (h) **Pemetaan UMKM dalam Upaya Pengentasan Kemiskinan dan Penyerapan Tenaga Kerja Menggunakan Algoritma K-Means**(Herwinda Kurniadewi, Rizal Abdul Hakim, Mohhamad Jajuli, Jajam Haerul Jaman, 2022),

pada penelitian ini Hasil evaluasi performa pengelompokan UMKM di Kabupaten Purwakarta dalam upaya pengentasan kemiskinan dan penyerapan tenaga kerja tahun 2021 menggunakan algoritma *K-Means* menggunakan *silhouette coefficient* pada penelitian ini pada 3 *cluster* yaitu dengan nilai index 0.45; visualisasi pemetaan menggunakan *library matplotlib.pyplot dan geopandas*, bahasa pemrograman *python* didapatkan hasil pemetaan untuk memudahkan dalam memahami hasil pengelompokan algoritma *K-Means*, nilai *silhouette coefficient siluet* yang diperoleh 0,45, hal tersebut menunjukkan *cluster* cukup baik;

- (i) **Pemetaan Daerah Berdasarkan Jenis Usaha UMKM Dengan Algoritma *K-Means* di Jawa Barat**(Hendrik, Teady Matius Surya Mulyana, 2023), pada penelitian ini mengelompokkan UMKM di Provinsi Jawa Barat untuk mengetahui nilai *k* yang optimal dari dataset yang digunakan, serta mengetahui tingkat akurasi dari setiap *cluster*, metode dan algoritma yang digunakan pada penelitian ini yaitu algoritma *K-Means* dan metode Elbow, akurasi pengklasteran dengan algoritma *K-Means* yaitu, 66,7% dengan data yang digunakan sebanyak 54 data, dan metode *K-Means* dapat mengklasterkan dataset usaha mikro, kecil, dan menengah (UMKM) di Jawa Barat dengan mendapatkan hasil *cluster* mulai dari C1, C2, C3, C4, C5 dengan jumlah anggota dari 18, 40, 44, 57, 3 anggota. Hasil perhitungan dari metode elbow didapatkan nilai *K* optimal dari dataset yaitu 5 dilihat dari grafik yang menunjukkan penurunan yang stabil pada *cluster* ke-5;
- (j) **Pengelompokan Bidang Usaha Terhadap Bantuan Produktif Usaha Mikro (BPUM) Berdasarkan Wilayah Deli Serdang Menggunakan Metode *Clustering K-Means***(Tiara Jelita, Relita Buaton, Magdalena Simanjuntak, 2023), pada penelitian ini mengelompokkan bidang di masing-masing usaha yang dimiliki masyarakat, karena tidak semua jenis bidang usaha yang ada di masyarakat akan mendapatkan bantuan di Kabupaten Deli Serdang Provinsi Sumatera Utara, dengan memanfaatkan proses data mining menggunakan metode *clustering* pengklasteran dapat menyelesaikan masalah pengelompokan bidang usaha yang dimiliki masyarakat dengan menghitung jarak menggunakan Euclidean Distance pada analisa dengan variabel yang digunakan, berdasarkan hasil uji coba sebanyak 1004 data, yang dilakukan dengan MATLAB diperoleh grup 1 terdapat 383 data, grup 2 terdapat 261 data dan grup 3 terdapat 360 data dan berdasarkan hasil uji coba dengan RapidMiner diperoleh grup 1 terdapat 371 data, grup 2 terdapat 281 data dan grup 3 terdapat 352 data,

Tabel 2.7 Tinjauan Pustaka

No	Nama Peneliti	Judul	Sumber	Kontribusi
1	Aisya Basri, Abdul Mubarak, Hairil Kurniadi Siradjuddin, Saiful Do. Abdullah	Penentuan Jumlah Klaster Terbaik Pada K-Means Dalam Melihat Pola Klustering Data Mahasiswa Yang Telah Lulus	JATI(Jurnal Jaringan dan Teknologi Informasi ), Vol. 3, No. 1, Juni 2023, hlm. 80-86 <a href="https://ejournal.unkhair.ac.id/index.php/jati/article/download/60/39/222">https://ejournal.unkhair.ac.id/index.php/jati/article/download/60/39/222</a>	Menghasilnya output 2 klaster (tepat waktu dan tidak tepat waktu) dan untuk menentukan jumlah klaster terbaik dengan representasi menggunakan grafik Elbow
2	Nursyifa, Resti Noor Fahmi	Implementasi Metode <i>K-Means Clustering</i> dalam Analisis Persebaran UMKM di Jawa Barat	Journal of Information System Vol.6, No.2, November Tahun 2021 <a href="http://publikasi.dinus.ac.id/index.php/joins/article/view/5310">http://publikasi.dinus.ac.id/index.php/joins/article/view/5310</a>	Hasil penelitian ini menunjukkan metode <i>K-Means</i> efektif dan menggunakan pengujian <i>Davies Bouldin Index</i> serta <i>silhouette coefficient</i>
3	Oding Herdiana, Shanti Maulani, Eryan Ahmad Firdaus	Strategi Pemasaran Produk Industri Kreatif Menggunakan Algoritma <i>K-Means Clustering</i> Berbasis Particle Swarm Optimization	Jurnal Nuansa Informatika Vol. 15, No. 2, Juli Tahun 2021 <a href="https://journal.uniku.ac.id/index.php/ilkom">https://journal.uniku.ac.id/index.php/ilkom</a>	Hasil dari penelitian ini menggunakan algoritma <i>K-Means</i> untuk proses <i>clustering</i> dan <i>continua</i> pada data UMKM dan <i>Particle Swarm Optimization</i> (PSO)
4	Wisti Aristika, Wirajaya Hartono	Menentukan Pengaruh Media Sosial Facebook Terhadap Usaha	Jurnal Ilmu Komputer dan Bisnis, Vol. 11, No. 11, Mei Tahun 2020	Menghasilkan dua kelompok pengguna media sosial facebook berpengaruh, dan sangat

No	Nama Peneliti	Judul	Sumber	Kontribusi
		Mikro kecil dan Menengah (UMKM) Di Kecamatan Pekanbaru	<a href="https://ojs.stmikdharmaapalariau.ac.id/index.php/jikb/article/view/38">https://ojs.stmikdharmaapalariau.ac.id/index.php/jikb/article/view/38</a>	berpengaruh. Hasil dari pengolahan data yang semakin mendekati angka nol, maka kemiripan data anggota <i>cluster</i> akan semakin baik
5	Puntoriza, Charista Fibriani	Analisis Persebaran UMKM Kota Malang Menggunakan <i>Cluster K-Means</i>	Journal of Information System Vol.5, No. 1, Mei Tahun 2020 <a href="https://scholar.archive.org/work/eoph7iw2abarpctmd2zljisjgm/access/wayback/http://publikasi.dinus.ac.id/index.php/joins/article/download/3469/pdf">https://scholar.archive.org/work/eoph7iw2abarpctmd2zljisjgm/access/wayback/http://publikasi.dinus.ac.id/index.php/joins/article/download/3469/pdf</a>	memudahkan pemerintah terkait dalam hal memilih peminjaman modal, menentukan potensi usaha dan menetapkan strategi pemasaran yang dilakukan dengan algoritma <i>K-Means cluster analysis</i>
6	Wahyu Sudrajat, Idahm Cholid, Johannes Petrus	Penerapan Algoritma <i>K-Means Clustering</i> untuk Pengelompokan UMKM Menggunakan Rapidminer	Jurnal JUPITER, Vol. 14 No. 1 Bulan April Tahun 2022 <a href="https://jurnal.polsri.ac.id/index.php/jupiter/article/view/4467">https://jurnal.polsri.ac.id/index.php/jupiter/article/view/4467</a>	Menghasilkan 3 <i>cluster</i> UMKM dengan menggunakan metode <i>K-Means</i> dan pengujian menggunakan <i>rapidminer</i>
7	Dona Marcelina, Annisa Kurnia, Terttiaavini	Analisis Klaster Kinerja Usaha Kecil dan Menengah Menggunakan Algoritma <i>K-Means Clustering</i>	Indonesian Journal of Machine Learning and Computer Science, Vol.3, No.2 Oktober Tahun 2023 <a href="https://journal.irpi.or.id/index.php/malcom/article/view/952">https://journal.irpi.or.id/index.php/malcom/article/view/952</a>	Menghasilkan 3 (tiga) <i>cluster</i> kelompok data yang saling berdekatan dengan inialisasi, yaitu UKM mandiri, UKM berkembang dan UKM binaan

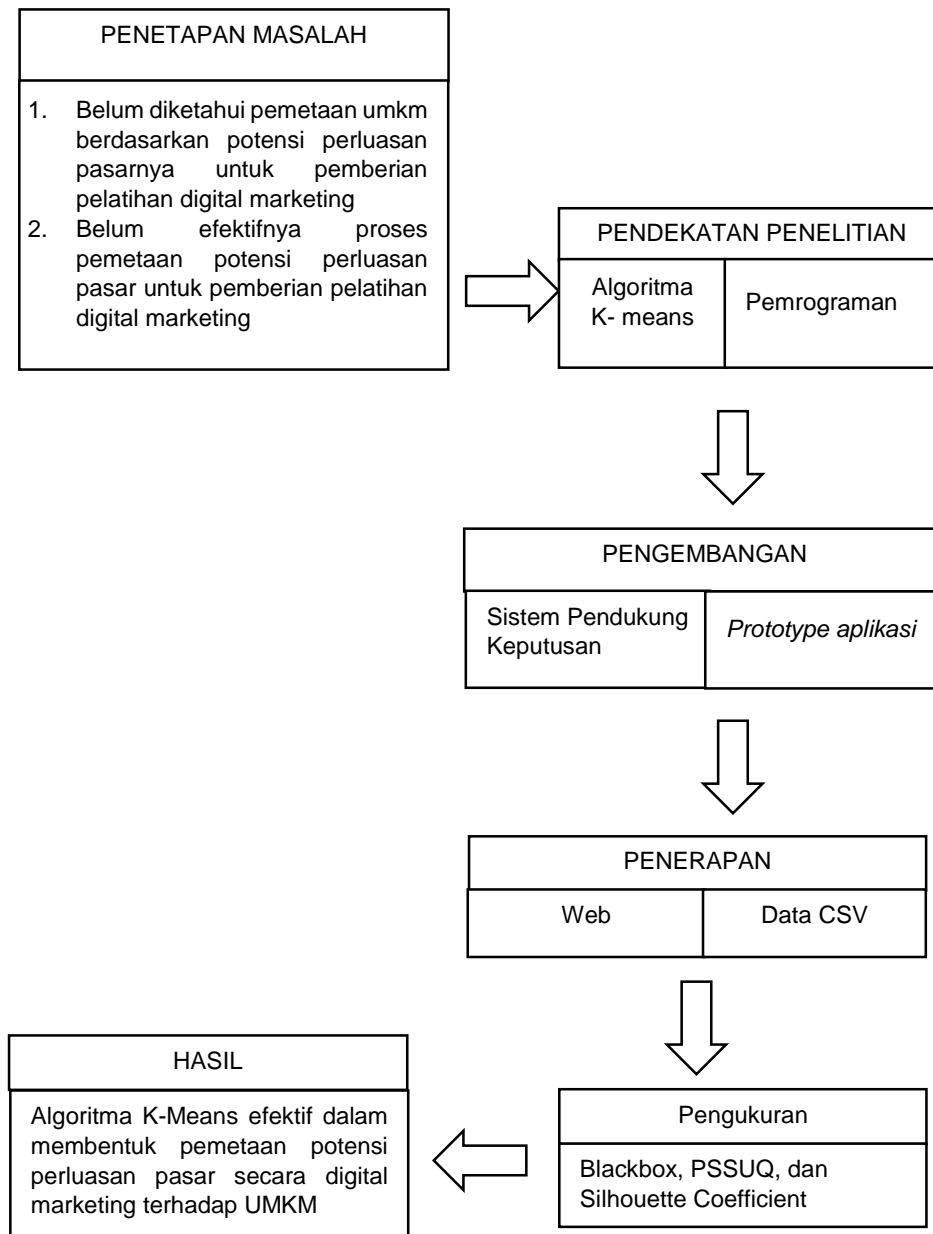
No	Nama Peneliti	Judul	Sumber	Kontribusi
8	Herwinda Kurniadewi, Rizal Abdul Hakim, Mohhama d Jajuli, Jajam Haerul Jaman	Pemetaan UMKM dalam Upaya Pengentasan Kemiskinan dan Penyerapan Tenaga Kerja	Journal of Applied Informatics and Coumputing(JAIC), Vol.6, No.2, Desember Tahun 2022 <a href="https://jurnal.polibatam.ac.id/index.php/JAIC/article/view/4227">https://jurnal.polibatam.ac.id/index.php/JAIC/article/view/4227</a>	Mengasilkan 3 <i>cluster</i> prioritas tinggi sebanyak 6 kecamatan, prioritas sedang sebanyak 8, dan prioritas rendah sebanyak 3 <i>cluster</i> yang menggunakan algoritma <i>K-Means</i> .
9	Hendrik, Teady Matius Surya Mulyana,	Pemetaan Daerah Berdasarkan Jenis Usaha UMKM Dengan Algoritma <i>K-Means</i> di Jawa Barat	Jurnal Teknologi Informasi, Komputer, dan Aplikasinya (JTIKA), Vol.5, No.2, September Tahun 2023 <a href="http://jtika.if.unram.ac.id/index.php/JTIKA/article/view/223">http://jtika.if.unram.ac.id/index.php/JTIKA/article/view/223</a>	Menghasilkan 3 <i>cluster</i> untuk mengetahui nilai k yang optimal dari dataset yang digunakan, serta mengetahui tingkat akurasi dari setiap klatser dengan menggunakan algoritma <i>K-Means</i> dan pengujian dengan metode <i>elbow</i>
10	Tiara Jelita, Relita Buaton, Magdalena Simanjuntak	Pengelompokan Bidang Usaha Terhadap Bantuan Produktif Usaha Mikro (BPUM) Berdasarkan Wilayah Deli Serdang Menggunakan Metode <i>Clustering K-Means</i>	Journal of Computer and Information Technology, Vol.3, No.2, Juli Tahun 2023 <a href="https://journal.fkpt.org/index.php/Explorer/article/view/783">https://journal.fkpt.org/index.php/Explorer/article/view/783</a>	Mengelompokkan data bidang usaha masyarakat dengan menggunakan metode <i>K-Means</i> dan pengujian menggunakan <i>tools Rapidminer</i>

Berdasarkan tinjauan pustaka yang sudah dipaparkan pada Tabel 2.7, didapatkan pengetahuan yang dijadikan rujukan dalam pelaksanaan penelitian ini. Rujukan penelitian ini berkontribusi dalam memberi pengetahuan ilmu mengenai pemetaan potensi perluasan pasar secara digital *marketing* terhadap UMKM. Penelitian yang dilakukan merupakan pengembangan dari penelitian yang ada sebelumnya. Pada penelitian ini, dilakukan pengembangan penelitian dengan

menggunakan metode *K-Means* dan di uji dengan *silhouette coefficient* pada usaha mikro, kecil, dan menengah untuk mengetahui intensitas perluasan pasar secara digital marketing.

### C. Kerangka Pemikiran

Kerangka pemikiran adalah skema untuk merangkai konsep pemikiran pada penelitian. Gambar 2.5 merupakan kerangka pemikiran pada penelitian ini;



Gambar 2.5 Kerangka Pemikiran



Keterangan kerangka pemikiran gambar 2.5 :

- (1) Permasalahan pada penelitian yaitu Belum diketahui pemetaan umkm berdasarkan potensi perluasan pasarnya untuk pemberian pelatihan digital marketing dan belum efektifnya proses pemetaan potensi perluasan pasar untuk pemberian pelatihan digital marketing;
- (2) Untuk menemukan solusi dari permasalahan, maka dilakukan pendekatan penelitian dengan menggunakan metode *K-Means* dan pemograman;
- (3) Pengembangan penelitian ini dilakukan dengan sistem pendukung keputusan dan membuat *prototyping* menggunakan bahasa pemograman *python streamlit*;
- (4) Penerapan penelitian dilakukan dengan menggunakan web dan data CSV;
- (5) Untuk menguji penelitian dan pengembangan yang dilakukan, maka dilakukan pengukuran dengan menggunakan *blackbox*, *PSSUQ*, dan *silhouette coefficient*;
- (6) Hasil dari penelitian dan pengembangan adalah terbukti bahwa menggunakan algoritma *K-Means* optimal dalam penentuan potensi perluasan pasar untuk pemberian pelatihan digital *marketing*.

#### **D. Hipotesis Penelitian**

Algoritma *K-Means* merupakan metode data *mining* yang digunakan untuk melakukan klusterisasi dengan mengelompokkan objek berdasarkan ketidaksamaan ataupun kesamaan dari karakteristiknya. Merujuk pada ketercapaian penelitian sebelumnya pada Penentuan Jumlah Kluster Terbaik Pada *K-Means* Dalam Melihat Pola Klustering Data Mahasiswa Yang Telah Lulus, metode *K-Means* mampu mengelompokkan data berdasarkan jumlah kluster optimal yang terdiri dari kluster tepat waktu dan kluster tidak tepat waktu. Berdasarkan pada pernyataan, belum diketahui pemetaan umkm berdasarkan potensi perluasan pasarnya untuk pemberian pelatihan digital *marketing* dan belum efektifnya proses pemetaan potensi perluasan pasar untuk pemberian pelatihan digital *marketing*, maka diperlukannya klusterisasi umkm untuk mengetahui potensi perluasan pasarnya dalam pemberian pelatihan digital *marketing* yang lebih efektif untuk memberikan informasi yang tepat pada pihak terkait dalam mengambil keputusan. Penerapan metode *K-Means* diduga dapat memberikan pemetaan potensi perluasan pasar secara digital marketing terhadap umkm.