

BAB II

LANDASAN TEORI

2.1 Penelitian Terdahulu

Penelitian terdahulu menjadi landasan untuk membandingkan, melengkapi, dan memperkuat penelitian yang sedang dilakukan. Pada subbab ini disajikan sepuluh penelitian terdahulu yang relevan dengan topik segmentasi profil risiko investor cryptocurrency menggunakan algoritma K-Means Clustering, perilaku investor mahasiswa, serta literasi keuangan digital. Penelitian terdahulu ini dirangkum dalam tabel komparatif untuk memudahkan analisis kesenjangan penelitian (research gap).

Tabel 2. 1. Penelitian Terdahulu

No	Judul, Author, Tahun Terbit	Metode/Algoritma	Hasil
1	Profiling investor behavior in the Malaysian derivatives market using K-means clustering, Tan et al., 2023 [1]	K-Means Clustering	Berhasil mengidentifikasi lima profil trader yang mendukung strategi manajemen risiko berbasis data.
2	Online risk-based portfolio allocation on subsets of crypto assets applying a prototype-based clustering algorithm, Lorenzo & Arroyo, 2023 [2]	Prototype-based clustering algorithm (K-Means)	Metode ini terbukti meningkatkan indikator profit dan risk-profit dibanding metode tradisional.
3	Maximising investors wealth in the cryptocurrency market: a behavioural perspective, Laryea et al., 2025 [3]	K-Means Clustering, Behavioral Scoring (CPT)	Klasterisasi aset berdasarkan karakteristik perilaku mampu meningkatkan kinerja portofolio kripto.
4	Diversified Crypto Assets Portfolio Optimization Using K-Means Clustering Algorithm And The	K-Means Clustering, Efficient Frontier	Model portofolio yang dibangun outperform (mengungguli) indeks pasar kripto utama.

No	Judul, Author, Tahun Terbit	Metode/Algoritma	Hasil
	Efficient Frontier, Setiawan & Hakim, 2023 [4]		
5	A K-means clustering model for analyzing the Bitcoin extreme value returns, Das et al., 2023 [5]	K-Means Clustering Model	Berhasil mengklasifikasi return ekstrem untuk membantu investor memahami risiko hari baik dan buruk di pasar.
6	Risk monitor system for blockchain-based security issuing system(Liu et al., 2025 [6]	K-Means Clustering, Sistem Berbasis VaR (Value at Risk)	Sistem berhasil mendeteksi pola volatilitas untuk peringatan dini risiko sistemik secara efektif di pasar kripto.
7	A systematic literature review of investor behavior in the cryptocurrency markets, Almeida & Gonçalves, 2023) [7]	Systematic Literature Review (SLR)	Faktor demografis, sosial, dan psikologis adalah pengaruh utama yang membentuk segmentasi perilaku investasi.
8	Pengaruh Afikasi Diri, Persepsi Risiko, dan Modal Minimal pada Minat Mahasiswa Berinvestasi Cryptocurrency, Wardhana & Putri, 2025 [8]	Analisis Kuantitatif / Studi Empiris	Self-efficacy (afikasi diri) dan modal berpengaruh positif, sedangkan persepsi risiko berpengaruh negatif pada minat investasi.
9	Determinants of cryptocurrency investment decisions, Study of students in Bali)(Rahyuda & Candradewi, 2023 [9]	Analisis Kuantitatif / Studi Empiris	Literasi keuangan, persepsi risiko, dan herding behavior memengaruhi keputusan investasi mahasiswa Bali secara signifikan.
10	Clustering Sukuk Using the K-Means Algorithm for Allocation of Investors Based on Investment Risk	K-Means Algorithm	Metode membedakan investor defensif hingga agresif secara akurat dan sangat dapat

No	Judul, Author, Tahun Terbit	Metode/Algoritma	Hasil
	Profile, Novika & Rahayu, 2024 [10]		diadaptasi untuk segmentasi mahasiswa di instrumen kripto.

Berdasarkan tinjauan sepuluh penelitian di atas, terdapat tiga kelompok temuan yang relevan dengan penelitian ini.

1. Kelompok pertama mencakup implementasi K-Means pada aset keuangan dan pasar kripto. Beberapa studi telah membuktikan bahwa algoritma K-Means sangat tangguh dalam menyeleksi dan mengalokasikan portofolio kripto yang mampu mengungguli kinerja indeks pasar tradisional [2], [4]. Pendekatan *clustering* ini terbukti akurat dalam mengklasifikasikan volatilitas return ekstrem pada Bitcoin, yang secara langsung membantu investor dalam memetakan risiko pada kondisi pasar yang anomali [5].
2. Kelompok kedua mencakup segmentasi perilaku dan profil risiko investor. Algoritma K-Means juga menunjukkan relevansi yang tinggi dalam memetakan profil perilaku pengguna, terbukti berhasil mengidentifikasi berbagai profil *trader* pada pasar derivatif untuk mendukung strategi manajemen risiko berbasis data [1]. Algoritma ini juga mumpuni ketika digunakan untuk mengalokasikan investor berdasarkan tingkat toleransi risiko pada instrumen keuangan lain seperti sukuk [10], serta integrasi skor perilaku ke dalam model *clustering* terbukti mampu memaksimalkan kinerja investasi portofolio [3].
3. Kelompok ketiga mencakup determinan investasi kripto pada demografi mahasiswa. Secara konseptual, faktor demografis, sosial, dan psikologis merupakan elemen fundamental yang membentuk pola perilaku investasi di bursa kripto [7]. Pada konteks lokal, studi empiris terhadap demografi mahasiswa di Indonesia menegaskan bahwa faktor self-efficacy dan modal awal berpengaruh positif terhadap minat investasi [8], sedangkan herding behavior, persepsi risiko, dan tingkat literasi keuangan menjadi determinan utama yang memengaruhi keputusan investasi mahasiswa [9].

Berdasarkan tinjauan ketiga kelompok tersebut, terdapat kesenjangan penelitian (*research gap*) yang cukup tajam. Mayoritas penelitian *machine learning* di ekosistem kripto saat ini masih terfokus pada optimasi teknikal dan alokasi aset [2], [4], [5], sedangkan literatur yang secara khusus mengintegrasikan metrik psikologis seperti FOMO ke dalam model *unsupervised learning* untuk menghasilkan segmentasi profil risiko mahasiswa berbasis data masih sangat terbatas. Penelitian ini dirancang untuk menjembatani celah tersebut.

2.2 Teori yang berkaitan

2.2.1 Sistem Informasi dan Data Mining

Sistem Informasi (SI) adalah kombinasi dari teknologi informasi dan aktivitas orang-orang yang menggunakan teknologi tersebut untuk mendukung operasi dan manajemen. Dalam konteks penelitian ini, SI berperan sebagai jembatan antara data mentah perilaku investor dengan pengambilan keputusan yang strategis. Pengolahan data dilakukan melalui Data Mining, yaitu proses penemuan pola tersembunyi, korelasi, dan anomali dalam set data besar untuk memprediksi hasil [12].

Penelitian ini menggunakan kerangka kerja CRISP-DM (*Cross-Industry Standard Process for Data Mining*) yang terdiri dari enam tahapan: *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modeling*, *Evaluation*, dan *Deployment*. Hal ini memastikan bahwa segmentasi profil risiko tidak hanya akurat secara teknis, tetapi juga relevan dengan kebutuhan industri keuangan [13].

2.2.2 Big Data dalam Sektor Keuangan

Fenomena *Big Data* dalam industri keuangan ditandai oleh konsep 4V [12]:

1. **Volume:** Jumlah data transaksi kripto yang dihasilkan setiap detik sangat masif.
2. **Velocity:** Kecepatan perubahan harga dan arus informasi di media sosial memengaruhi keputusan investor secara *real-time*.

3. Variety: Data yang diolah mencakup data terstruktur (riwayat transaksi) dan tidak terstruktur (hasil kuesioner teks).
4. Veracity: Tantangan dalam memastikan keakuratan data di tengah banyaknya informasi palsu (*hoax*) di pasar kripto.

2.2.3 Teknologi Blockchain dan Ekosistem Kripto

Blockchain adalah sebuah *Distributed Ledger Technology* (DLT) yang memungkinkan pencatatan transaksi secara permanen tanpa memerlukan perantara pusat (*trustless system*) [14]. Tiga konsep kunci dalam ekosistem blockchain yang relevan dengan penelitian ini adalah sebagai berikut.

Pertama, desentralisasi (*Decentralization*): data tidak disimpan di satu server tunggal, melainkan di ribuan *node* di seluruh dunia, sehingga meminimalisir risiko manipulasi data tunggal [14]. Kedua, *Smart Contracts*: protokol komputer yang memfasilitasi, memverifikasi, atau menegakkan negosiasi kontrak secara otomatis, yang memungkinkan lahirnya *Decentralized Finance* (DeFi). Ketiga, *Market Volatility Drivers*: selain faktor penawaran dan permintaan, volatilitas kripto dipengaruhi oleh penggunaan leverage yang tinggi oleh trader retail, yang dapat memicu likuidasi berantai saat harga berfluktuasi tipis [15].

2.2.4 Profil Risiko Investor: Definisi dan Tipologi

Profil risiko investor adalah representasi kuantitatif maupun kualitatif dari kemampuan dan kemauan seorang investor untuk menanggung risiko finansial dalam pengambilan keputusan investasi [19]. Dalam literatur manajemen investasi dan *behavioral finance*, profil risiko dikategorikan ke dalam empat tipologi utama. Penelitian ini mengintegrasikan prinsip *Modern Portfolio Theory* dari Markowitz [30] untuk menentukan komposisi rekomendasi aset kripto pada masing-masing profil berdasarkan kapasitas risiko dan orientasi waktunya [1], [10]:

1. Konservatif: Investor dengan toleransi risiko rendah yang memprioritaskan keamanan modal di atas potensi keuntungan. Ciri

khasnya adalah preferensi terhadap aset stabil jangka panjang dan sangat jarang terpengaruh oleh sentimen pasar jangka pendek [2]. Berdasarkan prinsip manajemen risiko [30], profil ini direkomendasikan untuk mengalokasikan mayoritas portofolionya pada aset kripto berkapitalisasi raksasa (*blue-chip*) yang volatilitasnya relatif lebih rendah, yaitu: 70% Bitcoin (BTC), 20% Ethereum (ETH), dan 10% Stablecoin.

2. Moderat: Investor dengan toleransi risiko menengah yang mencari keseimbangan antara pertumbuhan modal dan perlindungan risiko. Mereka umumnya mengkombinasikan strategi HODL untuk aset utama dengan *swing trading* terbatas untuk *altcoin* berkapitalisasi menengah. Secara teoretis [30], keseimbangan ini mendukung diversifikasi portofolio ke aset dengan rasio *risk-reward* menengah: 50% BTC, 30% ETH, dan 20% Altcoin Top-10.
3. Trader Agresif: Investor dengan toleransi risiko tinggi yang aktif memanfaatkan volatilitas pasar untuk keuntungan jangka pendek. Profil ini ditandai oleh pemahaman teknikal yang baik dan penggunaan instrumen berisiko [1]. Dengan kapasitas penyerapan risiko yang tinggi, profil ini cocok dengan eksposur volatilitas yang lebih luas [30], yakni: 30% BTC, 30% ETH, dan 40% Altcoin kapitalisasi menengah/kecil (*small/mid-cap*).
4. Spekulan: Investor dengan toleransi risiko sangat tinggi yang didominasi oleh bias emosional (FOMO) dan literasi teknikal sangat rendah. Kelompok ini cenderung berinvestasi pada aset berisiko sangat tinggi [7]. Sebagai deskripsi perilaku (bukan rekomendasi ideal), portofolio riil kelompok ini seringkali terdistorsi melawan prinsip diversifikasi yang sehat [30], umumnya dialokasikan pada 10% BTC, 20% ETH, dan 70% Memecoin/Microcap yang sangat rentan terhadap manipulasi pasar.

Keempat tipologi ini bukan sekadar kategorisasi arbitrer, melainkan berlandaskan pada teori *Behavioral Finance* dan telah digunakan secara konsisten dalam literatur segmentasi investor [1], [3], [10]. Penelitian ini

mengadopsi keempat tipologi tersebut sebagai label interpretasi hasil *clustering* K-Means.

2.2.5 Behavioral Finance dan Psikologi Investor Muda

Behavioral Finance adalah cabang ilmu keuangan yang mengintegrasikan temuan psikologi kognitif ke dalam teori keuangan standar, dengan argumen bahwa investor tidak selalu bertindak secara rasional (*homo economicus*) [7]. Dua teori utama dalam *behavioral finance* yang relevan dengan penelitian ini adalah sebagai berikut.

Prospect Theory menjelaskan dua bias utama dalam pengambilan keputusan finansial [16]. Pertama, *Risk Seeking in Losses*: investor cenderung mengambil risiko lebih besar saat dalam posisi merugi (berharap harga balik modal) dibandingkan saat sedang untung. Kedua, *Anchoring Bias*: kecenderungan investor untuk berpatokan pada harga tertinggi masa lalu (ATH) sebagai referensi harga masa depan, tanpa memperhatikan perubahan fundamental pasar.

Fear of Missing Out (FOMO) adalah fenomena psikologis yang dalam konteks investasi kripto merujuk pada kecemasan berlebihan untuk tidak melewatkan peluang keuntungan yang sedang viral di media sosial [22], [29]. Penelitian Melandri et al. [22] membuktikan bahwa FOMO berkorelasi positif dengan pengambilan keputusan investasi impulsif pada Generasi Z Indonesia, dan menjadi salah satu faktor utama yang mendorong mahasiswa menggunakan dana terlarang untuk berinvestasi. FOMO dalam penelitian ini diukur melalui dimensi *Emotional Bias* (X5) pada kuesioner.

Herding Behavior adalah kecenderungan investor untuk mengikuti keputusan kelompok mayoritas tanpa analisis independen [9]. Penelitian terdahulu membuktikan bahwa *herding behavior* merupakan determinan signifikan dalam keputusan investasi kripto mahasiswa di Indonesia [9]. Fenomena ini diperparah oleh ekosistem media sosial yang memungkinkan penyebaran sentimen pasar secara instan, sehingga investor dengan literasi rendah cenderung mengikuti tren tanpa pemahaman yang memadai.

2.2.6 *Dollar Cost Averaging (DCA)*

Dollar Cost Averaging (DCA) adalah strategi manajemen portofolio di mana seorang investor mengalokasikan sejumlah dana tetap secara rutin dan berkala ke dalam suatu aset, tanpa memperdulikan fluktuasi harga aset tersebut pada saat pembelian. Dalam konteks *behavioral finance*, strategi DCA sangat efektif untuk memitigasi bias emosional, kecemasan (*panic buying/selling*), dan risiko kesalahan penentuan waktu masuk pasar (*market timing*) [3]. Pada instrumen bervolatilitas ekstrem seperti cryptocurrency, DCA secara matematis menurunkan rata-rata biaya pembelian aset dari waktu ke waktu dan menstabilkan profil risiko investor [27]. Teori ini menjadi landasan fungsional untuk kalkulator simulasi DCA yang dikembangkan dalam sistem.

2.2.7 *Simulasi Monte Carlo dalam Manajemen Risiko*

Metode Monte Carlo adalah teknik pemodelan stokastik yang menggunakan sampel acak berulang (*random sampling*) untuk mensimulasikan berbagai kemungkinan hasil dari suatu proses yang memiliki variabel ketidakpastian tinggi, seperti pergerakan harga cryptocurrency [5]. Berbeda dengan proyeksi deterministik yang hanya menghasilkan satu garis pertumbuhan statis, simulasi Monte Carlo memetakan ratusan skenario (*price paths*) ke masa depan berdasarkan volatilitas dan ekspektasi imbal hasil (*return*) historis [30]. Dalam manajemen risiko portofolio modern [15], simulasi ini krusial untuk membantu investor memvisualisasikan rentang kemungkinan hasil terbaik (*best-case*) hingga terburuk (*worst-case*), sehingga ekspektasi keuntungan tidak semata-mata didorong oleh bias optimisme.

2.3 Tools/software yang digunakan

2.3.1 Algoritma Clustering: Tinjauan dan Perbandingan

Clustering adalah teknik *unsupervised machine learning* yang bertujuan mengelompokkan data ke dalam kluster berdasarkan kemiripan internal, tanpa memerlukan label kelas yang telah ditentukan sebelumnya [12]. Terdapat beberapa algoritma *clustering* utama yang umum digunakan dalam literatur, antara lain K-Means, DBSCAN (*Density-Based Spatial Clustering of Applications with Noise*), *Hierarchical Clustering*, dan *Gaussian Mixture Models* (GMM). Tabel 2.2 menyajikan perbandingan karakteristik keempat algoritma tersebut sebagai justifikasi pemilihan K-Means dalam penelitian ini.

Tabel 2. 2. Perbandingan Algoritma Clustering

Kriteria	K-Means	DBSCAN	Hierarchical Clustering	Gaussian Mixture Models
Kompleksitas komputasi	$O(nk^2)$ — rendah, efisien untuk data besar	$O(n \log n)$ — sedang	$O(n^2)$ — tinggi, tidak efisien untuk n besar	$O(n^3)$ per iterasi — tinggi
Jumlah kluster	Harus ditentukan di awal (k)	Otomatis dari densitas data	Dapat ditentukan setelah analisis dendrogram	Harus ditentukan di awal
Asumsi bentuk kluster	Spherical / convex	Kluster bentuk bebas	Fleksibel	Elipsoidal (Gaussian)
Sensitivitas terhadap outlier	Sensitif	Tidak sensitif (outlier = noise)	Sensitif	Cukup sensitif
Interpretabilitas hasil	Sangat tinggi (centroid = representasi profil)	Rendah-sedang	Tinggi (dendrogram)	Sedang (probabilistik)
Kesesuaian untuk data behavioral	Sangat sesuai	Kurang sesuai (data tidak ber-	Sesuai namun tidak efisien untuk	Sesuai namun over-kompleks

kuesioner (skala Likert)		densitas)	n=1.050	
Dukungan literatur pada profil risiko investor	Sangat kuat [1], [2], [3], [4], [5], [10]	Terbatas	Terbatas	Terbatas

Berdasarkan perbandingan pada Tabel 2.2, K-Means dipilih sebagai algoritma utama dengan empat justifikasi utama. Pertama, efisiensi komputasi: dengan kompleksitas $O(nkt)$, K-Means sangat efisien untuk dataset berukuran 1.050 responden dengan 5 dimensi, dibandingkan *Hierarchical Clustering* yang memiliki kompleksitas $O(n^2)$ sehingga kurang praktis untuk ukuran data tersebut. Kedua, interpretabilitas: centroid yang dihasilkan K-Means secara langsung merepresentasikan nilai rata-rata tiap dimensi perilaku dalam satu klaster, sehingga sangat mudah diinterpretasikan sebagai "profil rata-rata" investor pada segmen tersebut—sebuah keunggulan yang tidak dimiliki DBSCAN. Ketiga, kesesuaian data: data perilaku berbasis skala Likert 1–5 bersifat kontinu, terdistribusi merata, dan tidak memiliki outlier yang ekstrem, sehingga asumsi *spherical cluster* K-Means terpenuhi dengan baik. Keempat, validasi literatur yang kuat: K-Means telah terbukti efektif secara konsisten dalam segmentasi profil risiko investor pada penelitian-penelitian terdahulu yang relevan [1], [2], [3], [4], [10], memberikan basis komparatif yang solid bagi penelitian ini.

2.3.2 K-Means: Prosedur Kerja

K-Means merupakan algoritma *Unsupervised Learning* yang bertujuan untuk menemukan kelompok-kelompok (*clusters*) dalam data yang tidak memiliki label sebelumnya [17]. Algoritma ini dikategorikan sebagai teknik deskriptif untuk melakukan segmentasi populasi dalam Sistem Informasi [12]. Proses pengelompokan dilakukan melalui tahapan iteratif berikut:

1. Inisialisasi: Menentukan jumlah k (cluster) dan menempatkan k titik awal sebagai centroid. Penelitian ini menggunakan inisialisasi K-Means++

yang lebih deterministic dibandingkan inisialisasi acak murni, untuk mengurangi risiko *local minima* [18].

2. Assignment: Menghitung jarak setiap data poin ke setiap centroid menggunakan *Euclidean Distance* dan memasukkan poin tersebut ke klaster dengan centroid terdekat.
3. Update Centroid: Menghitung ulang rata-rata posisi semua poin dalam setiap klaster untuk memindahkan centroid ke posisi baru.
4. Konvergensi: Langkah 2 dan 3 diulang hingga posisi centroid tidak lagi berubah secara signifikan (perubahan WCSS < threshold).

2.3.3 Rumus Euclidean Distance

Algoritma K-Means sangat bergantung pada perhitungan jarak antar dimensi data (variabel perilaku). Rumus yang digunakan adalah *Euclidean Distance* [17]:

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

Rumus 2. 1. Euclidean Distance

Dalam penelitian ini, **x** merepresentasikan vektor data responden mahasiswa dengan 5 dimensi (X1–X5) dan **y** merepresentasikan vektor centroid cluster. Karena variabel memiliki rentang skala yang sama (skala Likert 1–5), standardisasi menggunakan Z-Score tetap dilakukan untuk memastikan tidak ada dimensi yang secara artifisial mendominasi perhitungan jarak [18].

2.3.4 Optimasi dan Evaluasi Jumlah Klaster

1. Elbow Method

Elbow Method digunakan untuk mencari jumlah klaster terbaik dengan melihat penurunan *Within-Cluster Sum of Squares* (WCSS) [19]. Rumus WCSS adalah:

$$WCSS = \sum_{i=1}^k \sum_{x \in C_i} ||x - \mu_i||^2$$

Rumus 2. 2. WCSS

Grafik akan menunjukkan titik "siku" (*elbow*) di mana penambahan jumlah kluster tidak lagi memberikan penurunan WCSS yang signifikan. Titik siku inilah yang dianggap sebagai jumlah kluster optimal [19].

2. Silhouette Coefficient

Selain *Elbow Method*, penelitian ini juga menggunakan *Silhouette Coefficient* untuk mengukur seberapa baik sebuah data poin cocok dengan klusternya dibandingkan dengan kluster lainnya [19].

Rumus *Silhouette Score* untuk satu titik data adalah:

$$s(i) = \max(a(i), b(i)) - a(i)$$

Rumus 2. 3. Silhouette Score

di mana $a(i)$ adalah rata-rata jarak titik i ke semua titik lain dalam kluster yang sama, dan $b(i)$ adalah rata-rata jarak titik i ke semua titik di kluster terdekat lainnya. Rentang nilai *Silhouette Score* adalah -1 hingga 1, di mana nilai mendekati 1 menunjukkan segmentasi yang sangat baik [19].

Interpretasi nilai *Silhouette Score* [19] disajikan pada Tabel 2.3.

Tabel 2. 3. Interpretasi Nilai Silhouette Score

Rentang Nilai	Interpretasi
0.71 – 1.00	Strong structure (struktur kluster sangat kuat)
0.51 – 0.70	Reasonable structure (struktur kluster cukup kuat)
0.26 – 0.50	Weak structure (struktur kluster lemah; perlu evaluasi tambahan)
< 0.25	No substantial structure (tidak ada struktur kluster yang berarti)

Nilai *Silhouette Score* pada data perilaku berbasis kuesioner (*behavioral survey data*) secara inheren cenderung lebih rendah dibandingkan data terstruktur seperti data transaksi finansial, disebabkan oleh overlap antar profil yang wajar terjadi pada

perilaku manusia [19]. Oleh karena itu, dalam penelitian berbasis perilaku, keputusan klaster optimal tidak ditentukan oleh satu metrik saja, melainkan menggunakan kombinasi dua metode: *Elbow Method* untuk melihat titik efisiensi WCSS dan *Silhouette Coefficient* untuk memilih nilai k dengan skor tertinggi di antara kandidat yang ada [23]. Dengan pendekatan kombinasi ini, k dengan skor *Silhouette* tertinggi dipilih sebagai optimal meskipun nilainya masuk kategori *weak structure*, selama diperkuat oleh konfirmasi *Elbow Method* dan justifikasi praktis dari teori behavioral finance.

2.3.5 Prototyping sebagai Metode Pengembangan Sistem

Prototyping adalah pendekatan metodologi pengembangan sistem informasi berbasis iterasi yang berfokus pada pembuatan model kerja (sampel) awal sistem untuk dievaluasi oleh pengguna sebelum sistem final dibangun sepenuhnya. Tujuan utama *prototyping* bukanlah menghasilkan sistem final secara langsung, melainkan untuk mengeksplorasi kebutuhan pengguna, mendemonstrasikan kelayakan sistem, dan memberikan umpan balik (*feedback*) mendalam mengenai desain antarmuka (UI) serta pengalaman pengguna (UX).

Karakteristik utama metode *prototyping* adalah keterlibatan pengguna yang sangat tinggi dan siklus hidup yang bersifat evolusioner. Prosesnya umumnya dimulai dengan analisis kebutuhan cepat, dilanjutkan dengan pembangunan model *prototype*, evaluasi oleh pengguna, dan perbaikan berulang berdasarkan masukan tersebut. Siklus ini terus berulang hingga pengguna merasa puas dan kebutuhan sistem terdefinisi secara jelas.

Pemilihan metodologi *Prototyping* dalam penelitian ini didasarkan pada tiga pertimbangan. Pertama, sistem yang dibangun memiliki komponen antarmuka UI/UX yang sangat bergantung pada umpan balik pengguna mahasiswa, sehingga evaluasi iteratif lebih efektif dibandingkan pengembangan linear. Kedua, logika pengelompokan K-Means memerlukan

iterasi eksperimental untuk menemukan jumlah kluster dan parameter yang paling representatif, yang sulit dilakukan pada metodologi *Waterfall* yang kaku. Ketiga, terjadi *pivot* besar pada tahap pengembangan, dari pendekatan *profiling-only* menjadi pendekatan *problem-solution*, yang hanya dapat difasilitasi secara efisien oleh metodologi *Prototyping*.

2.3.6 Standardisasi Data (Z-Score)

Variabel kuesioner dapat memiliki variansi sebaran data yang berbeda meskipun berada dalam skala yang sama, standardisasi data diwajibkan sebelum algoritma K-Means memproses perhitungan jarak Euclidean [12]. Standardisasi menggunakan Z-Score mentransformasi fitur data sehingga memiliki rata-rata (*mean*) 0 dan standar deviasi 1 [18]. Formula matematis untuk Z-Score adalah sebagai berikut:

$$z = \frac{(x - \mu)}{\sigma}$$

Rumus 2. 4. Standardisasi Z-Score

di mana z adalah nilai standar, x adalah nilai asli data, μ adalah rata-rata populasi, dan σ adalah standar deviasi.

2.4 Tools/Software yang Digunakan

2.4.1 Bahasa Pemrograman Python

Python digunakan sebagai bahasa pemrograman utama karena memiliki ekosistem yang sangat luas untuk Data Science dan Machine Learning. Python dipilih karena sintaksisnya yang intuitif serta dukungan library pihak ketiga yang memungkinkan pengolahan algoritma K-Means dilakukan secara efisien.

2.4.2 Visual Studio Code (IDE)

Sebagai lingkungan pengembangan terintegrasi (*Integrated Development Environment*), Visual Studio Code dengan ekstensi Python digunakan dalam penelitian ini. Selain itu, untuk eksplorasi data awal digunakan Jupyter Notebook yang memungkinkan penggabungan kode Python, teks naratif, persamaan matematika, dan visualisasi grafik di dalam satu dokumen interaktif. Langkah ini sangat membantu dalam proses *Exploratory Data Analysis* (EDA) serta dokumentasi tahap demi tahap pengklasteran.

2.4.3 Library Scikit-Learn

Library Scikit-Learn merupakan *tool* inti untuk implementasi *machine learning* [26]. Modul `sklearn.cluster` digunakan untuk menjalankan algoritma K-Means, modul `sklearn.preprocessing` untuk melakukan normalisasi data menggunakan `StandardScaler`, serta modul `sklearn.decomposition` untuk *Principal Component Analysis* (PCA). Penerapan ini krusial agar variabel dengan skala besar (seperti modal investasi) tidak mendominasi variabel dengan skala kecil (seperti tingkat FOMO 1-5) dalam perhitungan jarak Euclidean.

2.4.4 Library Pandas dan NumPy

Pandas digunakan untuk manipulasi data tabular (DataFrames) yang mempermudah proses pembersihan data (data cleaning), penanganan nilai yang hilang (missing values), dan konversi data dari Google Form (.csv) ke dalam format yang siap diolah. NumPy digunakan untuk operasi komputasi numerik tingkat tinggi, terutama dalam menangani struktur data multidimensional array yang dibutuhkan dalam perhitungan jarak antar titik data pada algoritma K-Means.

2.4.5 Library Plotly

Plotly digunakan untuk visualisasi data interaktif pada aplikasi web Streamlit. *Library* ini menggantikan kombinasi Matplotlib dan Seaborn untuk menghasilkan grafik yang lebih dinamis dan responsif terhadap interaksi pengguna. Plotly mendukung berbagai jenis *chart* yang diimplementasikan dalam aplikasi, antara lain *Radar Chart* untuk visualisasi 5 dimensi profil pengguna, *Scatter Plot* untuk pemetaan posisi *cluster* (PCA 2D), *Pie Chart* untuk melihat proporsi distribusi *cluster*, serta *Bar Chart* untuk menyajikan statistik responden.

Secara spesifik, grafik garis (*Line Chart*) dimanfaatkan untuk memvisualisasikan proyeksi pertumbuhan aset pada kalkulator *Dollar Cost Averaging* (DCA), serta memetakan ratusan skenario kemungkinan pergerakan harga (*price paths*) pada simulasi risiko Monte Carlo. Interaktivitas Plotly memungkinkan pengguna untuk melihat detail nilai proyeksi pada titik waktu tertentu hanya dengan mengarahkan kursor (*hover*). Terakhir, *Heatmap* digunakan untuk memvisualisasikan nilai *centroid* pada masing-masing *cluster* secara komparatif.

2.4.6 Streamlit Framework

Streamlit digunakan sebagai framework utama untuk membangun aplikasi web Crypto Risk Profiler. Streamlit memungkinkan pengembangan antarmuka web yang reactive hanya dengan kode Python, tanpa memerlukan HTML, CSS, dan JavaScript secara eksplisit (meskipun custom CSS digunakan untuk pastel theme). Versi yang digunakan adalah Streamlit 1.32.0 yang mendukung session state, custom components, dan multi-page navigation.

2.4.7 Library Requests dan CoinGecko API

Library Requests digunakan untuk berinteraksi dengan API eksternal CoinGecko untuk mendapatkan harga aset crypto secara real-time. CoinGecko API menyediakan endpoint gratis untuk mendapatkan data harga, perubahan persentase 24 jam, dan kapitalisasi pasar dari ribuan cryptocurrency. Sistem juga dilengkapi dengan caching mechanism (TTL 60 detik) untuk menghindari rate-limit dan fallback data jika API tidak tersedia.

2.4.8 Google Forms dan Microsoft Excel

Google Forms digunakan sebagai alat pengumpulan data primer secara daring untuk menjangkau responden mahasiswa. Microsoft Excel digunakan untuk pengecekan awal data (initial data screening) sebelum data diimpor ke dalam lingkungan Python.

2.4.9 Spesifikasi Perangkat Keras

Penelitian ini dilakukan menggunakan perangkat keras dengan spesifikasi: Sistem Operasi Windows 11 64-bit, Prosesor Intel Core i5 Generasi ke-10, dan Memori RAM 16 GB.