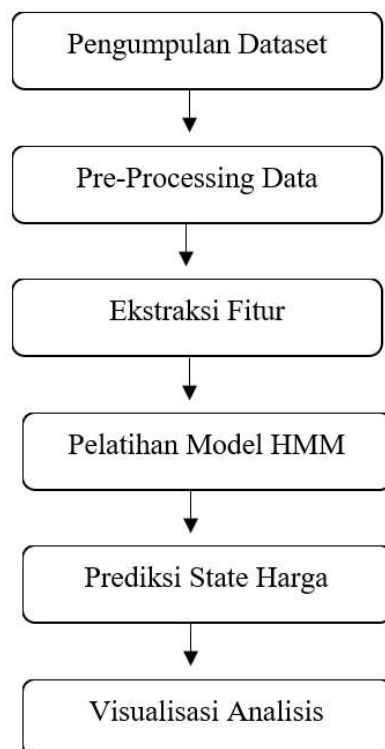


BAB III

METODE PENELITIAN

3.1 Desain Penelitian

Desain penelitian ini menggunakan pendekatan kuantitatif dengan metode analisis data historis untuk mengidentifikasi pola pergerakan harga *Bitcoin* berdasarkan teknik *machine learning Hidden Markov Model (HMM)*. Desain penelitian ini bertujuan untuk memberikan gambaran langkah-langkah yang akan diambil dalam penelitian penggunaan *Hidden Markov Model (HMM)* untuk memprediksi harga *Bitcoin*. Penelitian ini akan dilaksanakan melalui beberapa tahapan mulai dari pengumpulan data hingga visualisasi hasil prediksi. Penelitian



Gambar 3. 1 Desain Penelitian

Sumber: Data Peneliti, 2025

ini tidak bersifat eksperimental langsung terhadap objek nyata, melainkan bersifat rekonstruksi terhadap data historis yang tersedia secara publik.

3.1.1 Pengumpulan Dataset

Berikut ini merupakan dataset yang didapatkan dari *kaggle.com* sebanyak 1,048,576 baris, dengan jumlah 6 atribut sebagai berikut.

Timestamp	Open	High	Low	Close	Volume
1325412060.0	4.58	4.58	4.58	4.58	0.0
1325412120.0	4.58	4.58	4.58	4.58	0.0
1325412180.0	4.58	4.58	4.58	4.58	0.0
1325412240.0	4.58	4.58	4.58	4.58	0.0
1325412300.0	4.58	4.58	4.58	4.58	0.0
1325412360.0	4.58	4.58	4.58	4.58	0.0
1325412420.0	4.58	4.58	4.58	4.58	0.0
1325412480.0	4.58	4.58	4.58	4.58	0.0
1325412540.0	4.58	4.58	4.58	4.58	0.0
1325412600.0	4.58	4.58	4.58	4.58	0.0
1325412660.0	4.58	4.58	4.58	4.58	0.0
1325412720.0	4.58	4.58	4.58	4.58	0.0
1325412780.0	4.58	4.58	4.58	4.58	0.0
1325412840.0	4.58	4.58	4.58	4.58	0.0
1325412900.0	4.58	4.58	4.58	4.58	0.0
1325412960.0	4.58	4.58	4.58	4.58	0.0
1325413020.0	4.58	4.58	4.58	4.58	0.0
1325413080.0	4.58	4.58	4.58	4.58	0.0
1325413140.0	4.58	4.58	4.58	4.58	0.0
1325413200.0	4.58	4.58	4.58	4.58	0.0
1325413260.0	4.58	4.58	4.58	4.58	0.0
1325413320.0	4.58	4.58	4.58	4.58	0.0
1325413380.0	4.58	4.58	4.58	4.58	0.0
1325413440.0	4.58	4.58	4.58	4.58	0.0

Gambar 3. 2 Dataset

Sumber: Data Peneliti, 2025

Tahap awal dilakukan dengan mengumpulkan data historis harga *Bitcoin* dari sumber terbuka yang terpercaya. Dataset ini berisi informasi harga per

menit seperti harga pembukaan (*open*), tertinggi (*high*), terendah (*low*), penutupan (*close*), dan *volume* transaksi. Rentang data yang digunakan adalah dari tahun 2020 hingga 2025. Tujuan dari tahap ini adalah mendapatkan data numerik yang lengkap dan representatif untuk dilakukan analisis.

3.1.2 *Pre-Processing Data*

Setelah data terkumpul, tahap berikutnya adalah *pre-processing data*. Proses ini mencakup pembersihan dan transformasi data agar dapat digunakan dalam pelatihan model. Beberapa proses *pre-processing* yang dilakukan meliputi konversi format waktu dari *UNIX timestamp* ke format *datetime* yang mudah dibaca, serta penghitungan *Moving Average (MA)* selama 200 periode sebagai indikator teknikal untuk memperhalus fluktuasi harga. Selain itu, data yang memiliki nilai kosong (NaN) juga dihapus untuk memastikan model dapat berjalan tanpa gangguan.

3.1.3 *Ekstraksi Fitur*

Setelah preprocessing selesai, dilakukan tahap ekstraksi fitur. Fitur adalah komponen-komponen yang dianggap penting dan relevan dalam mempengaruhi hasil prediksi model. Dalam penelitian ini, fitur yang dipilih adalah harga penutupan (*Close*), nilai *Moving Average (MA)*, dan *volume* transaksi (*Volume*). Ketiga fitur ini dipilih karena mampu merepresentasikan kondisi pasar secara umum. Setelah diekstraksi, data fitur diubah menjadi array numerik yang akan digunakan dalam proses pelatihan model *HMM*.

3.1.4 Pelatihan Model *HMM*

Pada tahap ini, model *Hidden Markov* dilatih menggunakan fitur-fitur yang telah disiapkan. Model yang digunakan adalah *Gaussian Hidden Markov Model* dengan jumlah komponen tersembunyi (*hidden states*) sebanyak tiga. Proses pelatihan ini bertujuan agar model mampu mengenali pola tersembunyi dalam pergerakan harga berdasarkan distribusi probabilitas data. Parameter $n_iter=1000$ digunakan untuk memastikan proses training mencapai konvergensi. Model *HMM* ini akan mempelajari pola transisi antar kondisi pasar berdasarkan fitur input yang diberikan.

3.1.5 Prediksi *State* Harga

Setelah model *HMM* selesai dilatih, langkah selanjutnya adalah melakukan prediksi *state* terhadap data fitur. Model akan mengklasifikasikan setiap titik waktu pada data ke dalam salah satu dari tiga *state* tersembunyi yang telah dipelajari. Setiap *state* kemudian diinterpretasikan sebagai kondisi pasar, yaitu *bullish* (tren naik), *bearish* (tren turun), dan *sideways* (harga stagnan). Hasil prediksi ini memberikan gambaran yang lebih dalam terhadap pola pasar yang tidak terlihat secara langsung dari grafik harga biasa.

3.1.6 Visualisasi & Analisis

Setelah model *Hidden Markov Model (HMM)* selesai dilatih dan digunakan untuk memprediksi *state* tersembunyi dari data historis harga *Bitcoin*, tahap selanjutnya adalah melakukan visualisasi dan analisis terhadap hasil prediksi tersebut. Visualisasi dilakukan untuk membantu memahami bagaimana pola pergerakan harga *Bitcoin* terbagi ke dalam tiga kondisi pasar

yang diidentifikasi oleh model, yaitu *bullish* (tren naik), *bearish* (tren turun), dan *sideways* (harga stabil).

Visualisasi pertama berupa grafik harga *Bitcoin* secara keseluruhan dari tahun 2020 hingga 2025, dengan pewarnaan data berdasarkan masing-masing *state*. Hal ini bertujuan untuk memberikan gambaran umum mengenai bagaimana pergerakan harga diklasifikasikan oleh model ke dalam masing-masing kondisi pasar. Selanjutnya, dilakukan visualisasi tambahan berupa grafik dengan rentang waktu yang diperpendek, yaitu tahun 2021 hingga 2022. Visualisasi ini memberikan fokus yang lebih mendalam pada pergerakan harga dalam jangka waktu yang lebih singkat, serta memperjelas transisi antar *state* secara visual.

Selain grafik garis dan pewarnaan *state*, dilakukan pula analisis kuantitatif melalui bar chart yang menunjukkan distribusi jumlah data dalam masing-masing *state*. Analisis ini membantu melihat kondisi pasar mana yang paling sering terjadi selama periode penelitian. Untuk melengkapi hal tersebut, dilakukan analisis durasi rata-rata per *state*, yaitu pengukuran berapa lama (dalam satuan menit) sebuah kondisi pasar bertahan sebelum berpindah ke kondisi lainnya. Rata-rata durasi ini menjadi indikator penting untuk memahami stabilitas atau volatilitas dari masing-masing kondisi pasar.

Terakhir, hasil durasi *state* divisualisasikan dalam bentuk histogram untuk menggambarkan sebaran frekuensi lamanya setiap *state* berlangsung. Histogram ini menggambarkan durasi dalam satuan menit, dan dikelompokkan berdasarkan *state*, sehingga memudahkan dalam melihat seberapa sering

kondisi seperti *bullish* atau *bearish* terjadi hanya dalam waktu singkat atau cukup lama. Semua visualisasi ini digabungkan untuk membentuk narasi yang kuat mengenai perilaku harga *Bitcoin* berdasarkan klasifikasi tersembunyi dari model *HMM*.

3.2 Metode Pengumpulan Data

Dalam penelitian ini, metode pengumpulan data dilakukan melalui pendekatan pengumpulan data sekunder, yaitu data yang diperoleh dari sumber yang sudah tersedia dan dipublikasikan secara daring. Data yang digunakan merupakan data historis harga *Bitcoin* yang bersifat *time-series* dan dapat diakses dari platform penyedia data kripto secara terbuka seperti *kaggle.com*.

3.2.1 Jenis Data

Data yang dikumpulkan adalah data kuantitatif berupa: harga pembukaan (*Open Price*), harga tertinggi (*High Price*), harga terendah (*Low Price*), harga penutupan (*Close Price*), volume transaksi (*Volume*). Semua data bersifat historis dan dicatat dalam resolusi per satu menit (*1-min interval*), yang memungkinkan analisis granular terhadap pergerakan pasar dalam jangka pendek.

3.2.2 Metode Pengambilan

Data diperoleh dengan cara mengunduh file CSV yang telah disediakan oleh penyedia data. File ini berisi 1,048,576 entri data per menit dari tahun 2012 hingga 2025. File kemudian diolah menggunakan bahasa pemrograman *Python* dengan bantuan pustaka seperti *pandas* untuk manipulasi data, *numpy* untuk komputasi numerik, serta *matplotlib* untuk visualisasi.

Timestamp,Open,High,Low,Close,Volume
1325412060.0,4.58,4.58,4.58,4.58,0.0
1325412120.0,4.58,4.58,4.58,4.58,0.0
1325412180.0,4.58,4.58,4.58,4.58,0.0
1325412240.0,4.58,4.58,4.58,4.58,0.0
1325412300.0,4.58,4.58,4.58,4.58,0.0
1325412360.0,4.58,4.58,4.58,4.58,0.0
1325412420.0,4.58,4.58,4.58,4.58,0.0
1325412480.0,4.58,4.58,4.58,4.58,0.0
1325412540.0,4.58,4.58,4.58,4.58,0.0
1325412600.0,4.58,4.58,4.58,4.58,0.0
1325412660.0,4.58,4.58,4.58,4.58,0.0
1325412720.0,4.58,4.58,4.58,4.58,0.0
1325412780.0,4.58,4.58,4.58,4.58,0.0
1325412840.0,4.58,4.58,4.58,4.58,0.0
1325412900.0,4.58,4.58,4.58,4.58,0.0
1325412960.0,4.58,4.58,4.58,4.58,0.0
1325413020.0,4.58,4.58,4.58,4.58,0.0
1325413080.0,4.58,4.58,4.58,4.58,0.0
1325413140.0,4.58,4.58,4.58,4.58,0.0
1325413200.0,4.58,4.58,4.58,4.58,0.0
1325413260.0,4.58,4.58,4.58,4.58,0.0
1325413320.0,4.58,4.58,4.58,4.58,0.0
1325413380.0,4.58,4.58,4.58,4.58,0.0
1325413440.0,4.58,4.58,4.58,4.58,0.0

Gambar 3. 3 Dataset Bitcoin

Sumber: Data Peneliti, 2025

1331711940.0,5.3,5.3,5.3,5.3,0.0
1331712000.0,5.3,5.3,5.3,5.3,0.0
1331712060.0,5.3,5.3,5.3,5.3,0.0
1331712120.0,5.3,5.3,5.3,5.3,0.0
1331712180.0,5.3,5.3,5.3,5.3,0.0
1331712240.0,5.3,5.3,5.3,5.3,0.0
1331712300.0,5.3,5.3,5.3,5.3,0.0
1331712360.0,5.3,5.3,5.3,5.3,0.0
1331712420.0,5.3,5.3,5.3,5.3,0.0
1331712480.0,5.3,5.3,5.3,5.3,0.0
1331712540.0,5.3,5.3,5.3,5.3,0.0
1331712600.0,5.3,5.3,5.3,5.3,0.0
1331712660.0,5.3,5.3,5.3,5.3,0.0
1331712720.0,5.3,5.3,5.3,5.3,0.0
1331712780.0,5.3,5.3,5.3,5.3,0.0
1331712840.0,5.3,5.3,5.3,5.3,0.0
1331712900.0,5.3,5.3,5.3,5.3,0.0
1331712960.0,5.3,5.3,5.3,5.3,0.0
1331713020.0,5.3,5.3,5.3,5.3,0.0
1331713080.0,5.3,5.3,5.3,5.3,0.0
1331713140.0,5.3,5.3,5.3,5.3,0.0
1331713200.0,5.3,5.3,5.3,5.3,0.0
1331713260.0,5.3,5.3,5.3,5.3,0.0
1331713320.0,5.3,5.3,5.3,5.3,0.0
1331713380.0,5.3,5.3,5.3,5.3,0.0
1331713440.0,5.3,5.3,5.3,5.3,0.0

Gambar 3. 4 Dataset Bitcoin

Sumber: Data Peneliti, 2025

1331713500.0,5.3,5.3,5.3,5.3,0.0
1331713560.0,5.3,5.3,5.3,5.3,0.0
1331713620.0,5.3,5.3,5.3,5.3,0.0
1331713680.0,5.3,5.3,5.3,5.3,0.0
1331713740.0,5.3,5.3,5.3,5.3,0.0
1331713800.0,5.3,5.3,5.3,5.3,0.0
1331713860.0,5.3,5.3,5.3,5.3,0.0
1331713920.0,5.3,5.3,5.3,5.3,0.0
1331713980.0,5.3,5.3,5.3,5.3,0.0
1331714040.0,5.3,5.3,5.3,5.3,0.0
1331714100.0,5.3,5.3,5.3,5.3,0.0
1331714160.0,5.3,5.3,5.3,5.3,0.0
1331714220.0,5.3,5.3,5.3,5.3,0.0
1331714280.0,5.3,5.3,5.3,5.3,0.0
1331714340.0,5.3,5.3,5.3,5.3,0.0
1331714400.0,5.3,5.3,5.3,5.3,0.0
1331714460.0,5.3,5.3,5.3,5.3,0.0
1331714520.0,5.3,5.3,5.3,5.3,0.0
1331714580.0,5.3,5.3,5.3,5.3,0.0
1331714640.0,5.3,5.3,5.3,5.3,0.0
1331714700.0,5.3,5.3,5.3,5.3,0.0
1331714760.0,5.3,5.3,5.3,5.3,0.0
1331714820.0,5.3,5.3,5.3,5.3,0.0
1331714880.0,5.3,5.3,5.3,5.3,0.0
1331714940.0,5.3,5.3,5.3,5.3,0.0
1331715000.0,5.3,5.3,5.3,5.3,0.0

Gambar 3. 5 Dataset Bitcoin

Sumber: Data Peneliti, 2025

3.2.3 Alasan Menggunakan Data Sekunder

Penggunaan data sekunder dipilih karena beberapa alasan: Efisiensi waktu dan biaya, karena data sudah tersedia dan tidak memerlukan proses pengumpulan lapangan seperti survei atau wawancara. Relevansi tinggi, karena data berasal dari sumber kredibel yang telah digunakan dalam berbagai penelitian serupa. Kelengkapan data, data historis mencakup seluruh parameter yang dibutuhkan dalam proses analisis dan pemodelan.

3.2.4 Validasi dan Kebersihan Data

Meskipun data diperoleh dari sumber terpercaya, tetap dilakukan validasi awal berupa pengecekan terhadap missing values atau nilai kosong, penghapusan data duplikat, konversi format waktu dari *UNIX timestamp* ke format *datetime*, dan seleksi waktu agar hanya mencakup periode yang dibutuhkan. Langkah ini penting untuk memastikan bahwa data yang digunakan bersih, valid, dan layak digunakan dalam proses analisis.

3.2.5 Instrumen dan Tools

Dalam proses pengumpulan dan pemrosesan data, beberapa alat bantu yang digunakan adalah: Python: Sebagai bahasa pemrograman utama, *Jupyter Notebook*: Sebagai lingkungan kerja interaktif, Pustaka Python seperti: *pandas* untuk pemrosesan data, *matplotlib* untuk visualisasi, *HMMlearn* untuk pemodelan *Hidden Markov Model*.

3.3 Metode Penerapan Model

```

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from hmmlearn import hmm
import matplotlib.patches as mpatches
from itertools import groupby
import collections

np.random.seed(42)

data = pd.read_csv('btcusd_1-min_data.csv')
data['Timestamp'] = pd.to_datetime(data['Timestamp'], unit='s')
data = data[(data['Timestamp'] >= '2020-01-01') & (data['Timestamp'] < '2025-01-01')]
data['MA'] = data['Close'].rolling(window=200).mean()
features = data[['Close', 'MA', 'Volume']].dropna().values

model = hmm.GaussianHMM(n_components=3, covariance_type='full', n_iter=1000)
model.fit(features)

log_likelihood = model.score(features)
print(f"Log-Likelihood of the model: {log_likelihood:.2f}")

hidden_states = model.predict(features)

data = data.iloc[199:]
data['State'] = hidden_states

state_descriptions = {
    0: "Sideways (low volatility)",
    1: "Bullish (rising trend)",
    2: "Bearish (declining trend)",
}

state_colors = {
    0: 'yellow',
    1: 'green',
    2: 'red'
}

legend_handles = [
    mpatches.Patch(color='yellow', label='Sideways'),
    mpatches.Patch(color='green', label='Bullish'),
    mpatches.Patch(color='red', label='Bearish'),
    plt.Line2D([0], [0], color='blue', label='Close Price')
]

plt.figure(figsize=(15, 6))
plt.plot(data['Timestamp'], data['Close'], label='Close Price', color='blue', alpha=0.5)
plt.scatter(data['Timestamp'], data['Close'], c=data['State'].map(state_colors))
plt.title('Bitcoin Price and Hidden States (2020-2025)')
plt.xlabel('Date')
plt.ylabel('Price')
plt.legend(handles=legend_handles, loc='upper left')
plt.grid(True, alpha=0.3)
plt.tight_layout()
plt.show()

zoomed_data = data[(data['Timestamp'] >= '2021-01-01') & (data['Timestamp'] <= '2022-12-31')]

plt.figure(figsize=(15, 6))
plt.plot(zoomed_data['Timestamp'], zoomed_data['Close'], label='Close Price', color='blue', alpha=0.5)
plt.scatter(zoomed_data['Timestamp'], zoomed_data['Close'], c=zoomed_data['State'].map(state_colors))
plt.title('Bitcoin Price and Hidden States (2021-2022)')
plt.xlabel('Date')
plt.ylabel('Price')
plt.legend(handles=legend_handles, loc='upper left')
plt.grid(True, alpha=0.3)
plt.tight_layout()
plt.show()

state_counts = data['State'].value_counts().sort_index()
labels = [state_descriptions[i] for i in state_counts.index]
colors = [state_colors[i] for i in state_counts.index]

plt.figure(figsize=(8, 5))
plt.bar(labels, state_counts.values, color=colors)
plt.title('State Distribution Analysis')
plt.xlabel('State')
plt.ylabel('Number of Data Points (minutes)')
plt.xticks(rotation=40)
plt.grid(axis='y', alpha=0.3)
plt.tight_layout()
plt.show()

durasi_per_state = {0: [], 1: [], 2: []}
for k, g in groupby(hidden_states):
    durasi_per_state[k].append(len(list(g)))

print(f"Average Duration per State:")
for state, durations in durasi_per_state.items():
    avg_minutes = np.mean(durations)
    avg_days = avg_minutes / 1440
    print(f"{state_descriptions[state]}: {avg_minutes:.2f} minutes = {avg_days:.2f} days")

plt.figure(figsize=(10, 5))
for state, durations in durasi_per_state.items():
    plt.hist(durations, bins=30, alpha=0.6, label=state_descriptions[state], color=state_colors[state])

plt.title('Histogram of Duration per State (in Minutes)')
plt.xlabel('Duration (minutes)')
plt.ylabel('Frequency')
plt.legend()
plt.grid(True, alpha=0.3)
plt.xlim(0, 100)

```

Gambar 3. 6 Kode Hidden Markov Model

Sumber: Data Peneliti, 2025

Dalam penelitian ini, proses penerapan model program bertujuan untuk membangun model analisis pergerakan harga *Bitcoin* dengan pendekatan *machine learning* menggunakan algoritma *Hidden Markov Model (HMM)*. Metode yang digunakan mencakup beberapa tahapan penting, yaitu pengumpulan data, *pre-processing*, transformasi data, pemodelan algoritma, hingga visualisasi hasil klasifikasi. Setiap tahapan dirancang secara sistematis dan berkesinambungan untuk memastikan model mampu mengidentifikasi pola pasar yang tersembunyi. Berikut ini adalah tahapan metode perancangan yang dilakukan:

3.3.1 *Pre-processing* dan Pembersihan Data

Tahap awal dalam perancangan sistem adalah *pre-processing data*, yang bertujuan untuk memastikan bahwa data yang digunakan bersih dan relevan. Dataset yang digunakan diperoleh dari data historis harga *Bitcoin* per menit (*minute-level*) dengan format CSV yang memuat informasi seperti *timestamp*, *open*, *high*, *low*, *close*, dan *volume*. Data kemudian dikonversi ke format waktu yang dapat dibaca manusia dengan fungsi *pd.to_datetime*. Rentang data difokuskan pada periode antara tahun 2020 hingga 2025 untuk menjaga relevansi dan konsistensi analisis.

Selanjutnya, dilakukan perhitungan *moving average* (rata-rata pergerakan harga) dengan jendela waktu 200 menit, sebagai indikator teknikal tambahan. Proses ini ditunjukkan dalam kode berikut:

```
python
data = pd.read_csv('btcusd_1-min_data.csv')
data['Timestamp'] = pd.to_datetime(data['Timestamp'], unit='s')
data = data[(data['Timestamp'] >= '2020-01-01') & (data['Timestamp'] < '2025-01-01')]
data['MA'] = data['Close'].rolling(window=200).mean()
```

Gambar 3. 7 *Pre-Processing*

Sumber: Data Peneliti, 2025

`pd.read_csv('btcusd_1-min_data.csv')` memuat file historis BTC per-menit ke DataFrame *data*. Kolom *Timestamp* yang masih berupa detik UNIX diubah ke format waktu nyata dengan `pd.to_datetime(..., unit='s')`. Lalu dilakukan filter rentang waktu 2020–2024 (hingga sebelum 2025-01-01) agar analisis fokus pada periode penelitian. Berikutnya dibuat indikator teknikal MA200 lewat `data['Close'].rolling(window=200).mean()`. Karena MA butuh 200 titik pertama, baris awal akan bernilai NaN; oleh karena itu saat membentuk matriks fitur dilakukan `dropna()` agar model tidak menerima nilai kosong. Tiga fitur yang dipakai adalah *Close*, *MA*, dan *Volume*, lalu diubah menjadi array NumPy melalui `.values` untuk diberikan ke model.

Berikut adalah hasil penggalan data yang sudah melalui *pre-processing* dan pembersihan data, serta nilai *MA200* yang sudah dihitung.

	Timestamp	Open	High	Low	Close	Volume
4207278	2020-01-01 03:19:00	7163.01	7164.80	7158.95	7158.95	0.081318
4207279	2020-01-01 03:20:00	7164.06	7164.06	7159.00	7159.00	0.059923
4207280	2020-01-01 03:21:00	7159.06	7164.66	7159.06	7164.66	0.038415
4207281	2020-01-01 03:22:00	7164.16	7164.66	7161.29	7163.58	0.699111
4207282	2020-01-01 03:23:00	7163.41	7163.41	7158.92	7163.13	0.025354
4207283	2020-01-01 03:24:00	7162.43	7162.43	7161.82	7161.82	0.054821
4207284	2020-01-01 03:25:00	7158.00	7158.00	7158.00	7158.00	0.045017
4207285	2020-01-01 03:26:00	7158.10	7158.30	7158.10	7158.30	0.023618
4207286	2020-01-01 03:27:00	7163.54	7163.54	7160.05	7160.05	0.338841
4207287	2020-01-01 03:28:00	7163.54	7163.54	7159.77	7159.77	0.239165
4207288	2020-01-01 03:29:00	7163.54	7169.06	7159.77	7165.24	1.270694
4207289	2020-01-01 03:30:00	7160.87	7166.08	7160.87	7166.08	0.048273
4207290	2020-01-01 03:31:00	7160.97	7166.08	7160.97	7166.08	0.840610
4207291	2020-01-01 03:32:00	7166.08	7166.08	7166.08	7166.08	0.000000
4207292	2020-01-01 03:33:00	7166.08	7167.87	7161.09	7167.87	0.195471
4207293	2020-01-01 03:34:00	7165.24	7168.89	7165.24	7168.89	0.451738
4207294	2020-01-01 03:35:00	7168.89	7171.45	7165.24	7171.45	0.376251
4207295	2020-01-01 03:36:00	7167.00	7172.47	7167.00	7172.47	0.512203
4207296	2020-01-01 03:37:00	7167.10	7173.24	7167.10	7173.24	2.172084
4207297	2020-01-01 03:38:00	7173.24	7175.00	7168.00	7173.71	0.859464
4207298	2020-01-01 03:39:00	7168.01	7173.71	7168.01	7173.71	0.139117
4207299	2020-01-01 03:40:00	7173.45	7173.45	7167.00	7167.00	0.773022
4207300	2020-01-01 03:41:00	7164.60	7172.52	7161.73	7172.51	0.070794
4207301	2020-01-01 03:42:00	7164.60	7164.60	7161.59	7161.59	0.015509
4207302	2020-01-01 03:43:00	7162.03	7169.96	7162.03	7169.96	0.088079
4207303	2020-01-01 03:44:00	7169.58	7173.16	7169.58	7173.16	0.672025
4207304	2020-01-01 03:45:00	7162.78	7170.62	7162.78	7163.69	0.101086
4207305	2020-01-01 03:46:00	7170.62	7170.62	7162.89	7164.60	0.619227
4207306	2020-01-01 03:47:00	7165.60	7172.41	7165.04	7172.41	2.212680
4207307	2020-01-01 03:48:00	7166.14	7166.24	7166.14	7166.24	0.036610
4207308	2020-01-01 03:49:00	7166.24	7172.80	7166.24	7172.80	0.259396
4207309	2020-01-01 03:50:00	7172.90	7175.00	7166.26	7173.26	3.026609
4207310	2020-01-01 03:51:00	7175.00	7179.51	7173.26	7179.51	2.439208
4207311	2020-01-01 03:52:00	7179.51	7186.35	7179.51	7186.35	1.454733
4207312	2020-01-01 03:53:00	7180.00	7180.31	7176.16	7176.16	6.402867
4207313	2020-01-01 03:54:00	7176.16	7182.22	7173.58	7182.22	0.347226
4207314	2020-01-01 03:55:00	7180.31	7180.31	7174.01	7174.01	3.051187
4207315	2020-01-01 03:56:00	7174.01	7180.81	7174.01	7180.81	1.790169
4207316	2020-01-01 03:57:00	7175.32	7175.32	7175.32	7175.32	0.007563
4207317	2020-01-01 03:58:00	7175.53	7178.88	7174.11	7178.88	0.093340
4207318	2020-01-01 03:59:00	7174.01	7179.93	7171.33	7171.33	6.346949
4207319	2020-01-01 04:00:00	7175.39	7175.39	7171.23	7175.39	0.518345
4207320	2020-01-01 04:01:00	7175.39	7175.39	7175.39	7175.39	0.115856
4207321	2020-01-01 04:02:00	7172.96	7178.31	7172.96	7173.86	0.860977
4207322	2020-01-01 04:03:00	7170.76	7171.94	7166.25	7166.25	7.476673
4207323	2020-01-01 04:04:00	7170.44	7170.44	7170.44	7170.44	0.025002
4207324	2020-01-01 04:05:00	7170.44	7170.44	7167.10	7167.10	0.046825
4207325	2020-01-01 04:06:00	7167.10	7167.10	7165.51	7165.51	0.043550

Gambar 3. 8 Hasil Data Pre-Processing

Sumber: Data Peneliti, 2025

	MA
4207278	7148.88665
4207279	7148.96095
4207280	7149.11645
4207281	7149.24860
4207282	7149.41225
4207283	7149.56935
4207284	7149.70530
4207285	7149.84595
4207286	7149.99475
4207287	7150.13495
4207288	7150.28995
4207289	7150.43285
4207290	7150.61090
4207291	7150.75895
4207292	7150.94680
4207293	7151.13975
4207294	7151.27015
4207295	7151.39670
4207296	7151.55195
4207297	7151.70955
4207298	7151.84805
4207299	7151.95250
4207300	7152.09205
4207301	7152.17650
4207302	7152.27015
4207303	7152.40575
4207304	7152.48175
4207305	7152.54715
4207306	7152.64175
4207307	7152.69800
4207308	7152.81645
4207309	7152.91155
4207310	7153.03905
4207311	7153.20085
4207312	7153.34140
4207313	7153.53635
4207314	7153.66250
4207315	7153.85460
4207316	7154.00815
4207317	7154.14210
4207318	7154.27230
4207319	7154.40450
4207320	7154.58000
4207321	7154.73470
4207322	7154.82590
4207323	7154.97545
4207324	7155.09095
4207325	7155.19850
4207326	7155.30960

Gambar 3. 9 Hasil Analisis *MA200*

Sumber: Data Peneliti, 2025

3.3.2 Integrasi dan Seleksi Fitur

Setelah proses *cleaning*, langkah selanjutnya adalah integrasi dan pemilihan fitur. Dalam penelitian ini, fitur yang digunakan untuk melatih model adalah harga penutupan (*close*), nilai *moving average*, dan volume perdagangan. Fitur-fitur ini dipilih karena memiliki keterkaitan langsung terhadap pergerakan tren harga di pasar kripto. Seluruh fitur digabung dalam bentuk array numerik dan baris dengan nilai kosong (*missing value*) dihilangkan menggunakan *.dropna()* agar tidak mengganggu proses pelatihan model. Hasilnya disimpan dalam variabel *features*:



```
python
features = data[['Close', 'MA', 'Volume']].dropna().values
```

Gambar 3. 10 Kode Integrasi Seleksi Fitur

Sumber: Data Peneliti, 2025

3.3.3 Transformasi Data dan Pelatihan Model *HMM*

Tahapan inti dari proses ini adalah transformasi data melalui pelatihan menggunakan algoritma *Gaussian Hidden Markov Model (HMM)*. Model dilatih untuk mengidentifikasi pola tersembunyi (*hidden states*) dari data berdasarkan distribusi *Gaussian multivariat*. Jumlah komponen (*hidden states*) ditentukan sebanyak tiga, yang merepresentasikan tiga kondisi pasar utama: *bullish*, *bearish*, dan *sideways*. Proses pelatihan model dilakukan dengan memanggil metode *.fit()* dan hasil klasifikasi disimpan dalam variabel *hidden_states*. Implementasi kode sebagai berikut:

```
python
model = hmm.GaussianHMM(n_components=3, covariance_type="full", n_iter=1000)
model.fit(features)
hidden_states = model.predict(features)
```

Gambar 3. 11 Pelatihan Model HMM

Sumber: Data Peneliti, 2025

`hmm.GaussianHMM(n_components=3, covariance_type="full", n_iter=1000)` membentuk model HMM dengan 3 state tersembunyi (ditafsirkan sebagai sideways, bullish, bearish). `covariance_type="full"` berarti setiap state memodelkan observasi dengan multivariate Gaussian yang memiliki matriks kovarian penuh (relasi antarfitur tidak dipaksa independen). `n_iter=1000` adalah batas maksimum iterasi algoritma EM (Baum–Welch) saat mencari parameter terbaik: peluang awal tiap state, matriks transisi antar state, serta mean dan kovarian Gaussian untuk masing-masing state. `model.fit(features)` menjalankan proses pelatihan tersebut di seluruh urutan data fitur.

Setelah proses pelatihan selesai, hasil prediksi dikembalikan ke dalam struktur *DataFrame* dan dikaitkan dengan data historis *Bitcoin*:

```
python
data = data.iloc[199:] # Menghapus baris awal akibat rolling window
data['State'] = hidden_states
```

Gambar 3. 12 Struktur *DataFrame*

Sumber: Data Peneliti, 2025

3.3.4 Visualisasi dan Interpretasi *Hidden States*

Setelah model berhasil mengklasifikasikan data ke dalam *hidden states*, tahap berikutnya adalah visualisasi hasil model untuk memberikan pemahaman yang lebih intuitif terhadap pola pergerakan pasar. Data ditampilkan dalam bentuk grafik garis harga *Bitcoin* yang diwarnai berdasarkan hasil klasifikasi *state*. Selain itu, dibuat juga tampilan *zoomed-in* untuk periode 2021–2022, bar chart distribusi jumlah data tiap *state*, serta histogram durasi tiap *state*. Visualisasi ini mempermudah interpretasi model dan memberikan gambaran kondisi pasar berdasarkan waktu dan tren. Berikut salah satu contoh kode visualisasinya:

```
python
plt.plot(data['Timestamp'], data['Close'], label='Close Price', color='blue', alpha=0.5)
plt.scatter(data['Timestamp'], data['Close'], c=data['State'].map(state_colors))
plt.title('Bitcoin Price and Hidden States (2020-2025)')
```

Gambar 3. 13 Visualisasi *Hidden States*

Sumber: Data Peneliti, 2025

3.3.5 Analisis Statistik Durasi *Hidden State*

Untuk memberikan pemahaman lebih mendalam terhadap pola durasi tiap keadaan pasar, dilakukan analisis statistik terhadap lamanya data berada dalam suatu *state*. Proses ini dilakukan dengan mengelompokkan *hidden states* secara berurutan dan menghitung durasinya dalam satuan menit. Rata-rata durasi dihitung dan dikonversikan ke satuan hari, sehingga diperoleh informasi berapa lama pasar biasanya berada dalam kondisi *bullish*, *bearish*, atau *sideways*. Ini

berguna untuk memahami stabilitas atau volatilitas masing-masing kondisi pasar. Potongan kodenya adalah sebagai berikut:

```
python
durasi_per_state = {0: [], 1: [], 2: []}
for k, g in groupby(hidden_states):
    durasi_per_state[k].append(len(list(g)))

for state, durations in durasi_per_state.items():
    avg_minutes = np.mean(durations)
    avg_days = avg_minutes / 1440
    print(f"{state_descriptions[state]}: {avg_minutes:.2f} minutes ≈ {avg_days:.2f} days")
```

Gambar 3. 14 Analisis Statistik Durasi

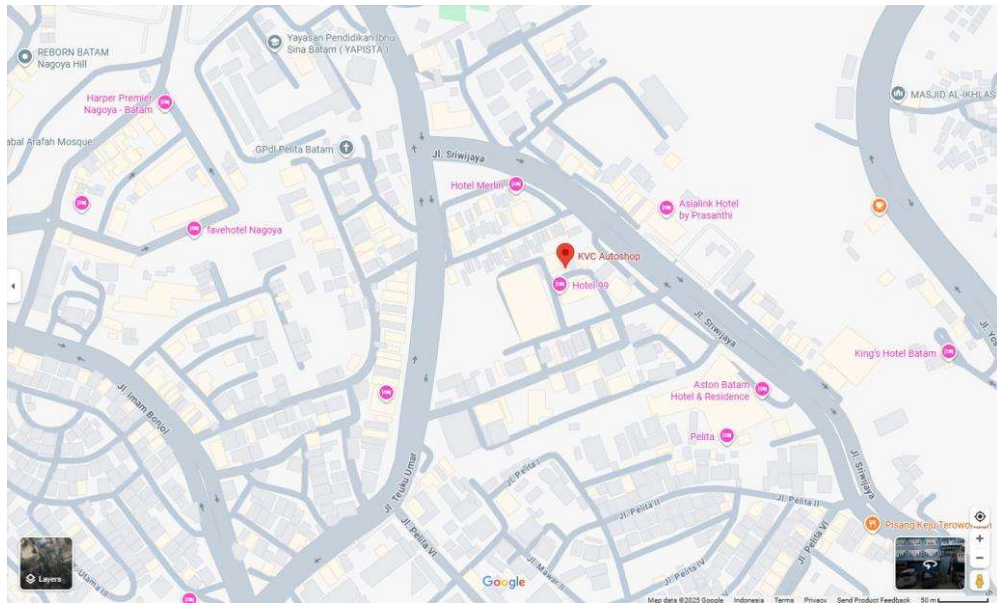
Sumber: Data Peneliti, 2025

Untuk mengetahui berapa lama suatu rezim biasanya bertahan, digunakan *groupby(hidden_states)* dari *itertools*—bukan *groupby* milik *pandas*. Fungsi ini mengelompokkan segmen berurutan yang memiliki label state sama (mis. 1,1,1 lalu ganti ke 0,0, dst.). Panjang setiap segmen (*len(list(g))*) adalah durasi dalam menit dari state tersebut sebelum berganti. Durasi-durasi ini dikumpulkan ke *durasi_per_state* (dictionary of lists). Lalu dihitung rata-rata menit per state dengan *np.mean(durations)* dan juga dikonversi ke hari dengan membagi 1440. Hasilnya dicetak agar Anda bisa melaporkan “rata-rata bullish bertahan X menit, bearish Y menit, dst.”

3.4 Lokasi dan Jadwal Penelitian

Penelitian dilakukan secara daring menggunakan perangkat pribadi dan *software open-source*. Proses pengumpulan data, pelatihan model, dan analisis dilakukan dari kediaman peneliti atau tempat yang mendukung aktivitas riset yaitu

Komp. Pelita Square, Jl. Sriwijaya Blk. C No.1, Kp. Pelita, Kec. Lubuk Baja, Kota Batam, Kepulauan Riau 29444.



Gambar 3. 15 Lokasi Penelitian

Sumber: Data Peneliti, 2025

No	Jenis Kegiatan	Waktu Pelaksanaan																			
		Februari 2025				Maret 2025				April 2025				Mei 2025				Juni 2025			
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
1	Menyempurnakan Bab 1 sampai Bab 3																				
2	Mengajukan surat keterangan penelitian																				
3	Mengumpulkan data penelitian																				
4	Analisis hasil pengolahan data penelitian																				
5	Mempersiapkan naskah publikasi																				
6	Publikasi																				
7	Penyempurnaan laporan skripsi																				
8	Pengumpulan softcover, upload naskah publikasi ke SIAM																				

Tabel 3. 1 Waktu Penelitian

Sumber: Data Peneliti, 2025