

BAB II

TINJAUAN PUSTAKA

2.1 Teori Dasar

2.1.1 Teori Probabilitas dan Statistika

Teori probabilitas dan statistika merupakan fondasi utama dalam analisis data dan pengambilan keputusan di bawah ketidakpastian. Probabilitas digunakan untuk memodelkan ketidakpastian dan variabilitas dalam fenomena acak, sementara statistika menyediakan alat untuk menganalisis data dan membuat inferensi tentang populasi berdasarkan sampel.

Dalam konteks *machine learning*, teori probabilitas memainkan peran penting dalam pengembangan model prediktif. Misalnya, pendekatan *algorithmic probability* telah digunakan untuk meningkatkan akurasi klasifikasi dalam kondisi data pelatihan yang terbatas (Dingle et al., 2022). Selain itu, konsep *Optimal Transport* telah diterapkan sebagai kerangka probabilistik untuk membandingkan dan memanipulasi distribusi probabilitas dalam berbagai tugas *machine learning*, termasuk *generative modeling* dan *transfer learning* (Montesuma et al., 2023).

Distribusi probabilitas, baik diskrit maupun kontinu, digunakan untuk memodelkan berbagai jenis data. Misalnya, distribusi normal sering digunakan untuk data yang bersifat kontinu dan simetris, sementara distribusi binomial digunakan untuk data diskrit dengan dua kemungkinan hasil. Pemahaman

tentang distribusi ini penting dalam analisis statistik dan pengembangan model prediktif.

Inferensi statistik memungkinkan peneliti untuk membuat kesimpulan tentang populasi berdasarkan data sampel. Dalam era data besar, inferensi statistik telah berkembang untuk menangani data berdimensi tinggi, dengan pendekatan seperti reduksi dimensi dan estimasi parameter yang efisien (Ayyala, 2019). Selain itu, pendekatan *Bayesian* telah menjadi populer dalam analisis data, memungkinkan integrasi informasi prior dan data baru untuk memperbarui estimasi parameter secara dinamis.

Dalam praktiknya, metode statistik digunakan dalam berbagai bidang, termasuk genetika statistik, di mana meta-analisis statistik memungkinkan identifikasi varian genetik yang terkait dengan berbagai penyakit kompleks (Lin, 2023). Selain itu, model seperti Matérn telah digunakan dalam statistik spasial untuk estimasi dan prediksi, serta dalam *machine learning* untuk tugas-tugas seperti regresi dan klasifikasi (Porcu et al., 2023).

2.1.2 Model Markov

Proses Markov adalah jenis proses stokastik yang memiliki properti khusus di mana masa depan hanya bergantung pada keadaan saat ini dan tidak bergantung pada bagaimana keadaan tersebut dicapai (Grewal et al., 2019). Dalam istilah yang lebih formal, sebuah proses stokastik $(X_t)_{t \geq 0}$ adalah Proses Markov jika:

$$P(x_{t+1} = x | x_t = x_t, x_{t-1} = x_{t-1}, \dots, x_0 = x_0) = P(x_{t+1} = x | x_t = x_t)$$

Rumus 2. 1 Rumus Model Markov

Sumber: *ScienceDirect.com*

Keterangan.

x_t : variabel acak yang merepresentasikan keadaan (*state*) pada waktu t .

P : probabilitas terjadinya suatu peristiwa.

Bagian kiri rumus : peluang keadaan berikutnya (x_{t+1}) bergantung pada semua keadaan sebelumnya ($x_t, x_{t-1}, \dots, x_0$).

Bagian kanan rumus : peluang keadaan berikutnya (x_{t+1}) cukup bergantung pada keadaan saat ini (x_t) saja.

Proses ini dinamakan setelah Andrey Markov, seorang matematikawan Rusia yang pertama kali memperkenalkan konsep ini pada awal abad ke-20.

Properti utama dari proses Markov adalah sifat tanpa memori, yang menyederhanakan banyak analisis matematis karena masa depan proses hanya tergantung pada keadaan saat ini (Grewal et al., 2019)

1. Properti Markov (*Markov Property*)

Properti Markov menunjukkan bahwa probabilitas masa depan hanya bergantung pada keadaan saat ini, bukan pada urutan kejadian yang membawa sistem ke keadaan tersebut. Ini sering disebut sebagai properti tanpa memori.

2. Homogenitas Waktu (*Time Homogeneity*)

Dalam banyak kasus, probabilitas transisi dari satu keadaan ke keadaan lain tidak bergantung pada waktu. Ini berarti bahwa $P(X_{t+1} = j | X_t = i)$ adalah konstan untuk semua t .

3. Transisi Stasioner (*Stationary Transition Probabilities*)

Jika proses Markov adalah stasioner, maka probabilitas transisi dari satu keadaan ke keadaan lain tetap konstan dari waktu ke waktu.

Model Markov memungkinkan analisis yang lebih terstruktur dan dapat diandalkan dari proses stokastik dalam keuangan, membantu dalam memahami dinamika pasar dan risiko terkait (Haryanto, 2022). Model Markov digunakan dalam berbagai bidang, termasuk:

1. Ekonomi dan Keuangan

Model Markov digunakan untuk memodelkan harga aset, suku bunga, dan risiko kredit. Contoh yang terkenal adalah model rantai Markov untuk harga saham.

2. Pemodelan Cuaca

Proses Markov digunakan untuk memodelkan perubahan cuaca, di mana kondisi cuaca saat ini digunakan untuk memprediksi kondisi cuaca masa depan.

3. Pengolahan Sinyal dan Komunikasi

Proses Markov digunakan dalam pemodelan dan prediksi sinyal serta dalam algoritma deteksi kesalahan.

4. Bioinformatika

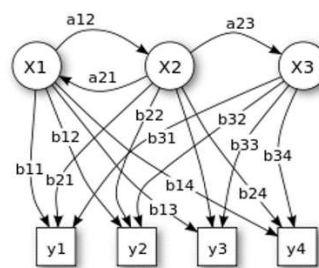
Proses Markov digunakan untuk memodelkan urutan DNA dan protein, memprediksi struktur sekunder RNA, dan dalam analisis genomik.

5. Pengolahan Bahasa Alami

Dalam NLP, model rantai Markov digunakan untuk tugas-tugas seperti penandaan bagian dari ucapan dan pengenalan bicara.

2.1.3 *Hidden Markov Model (HMM)*

Hidden Markov Model



Gambar 2. 1 *Hidden Markov Model*

Sumber: *ScienceDirect.com*

HMM adalah alat yang kuat untuk memodelkan sistem stokastik dengan struktur tersembunyi, memungkinkan analisis dan prediksi yang lebih efektif dalam berbagai aplikasi, dari pengenalan bicara hingga bioinformatika. (Mor et al., 2021) *HMM* didefinisikan oleh lima komponen utama:

2.1.3.1 Set Keadaan Tersembunyi (*Hidden States*)

Kumpulan keadaan yang tidak langsung dapat diamati tetapi menentukan proses yang menghasilkan observasi. Misalnya, dalam pengenalan ucapan, *hidden states* dapat berupa fonem-fonem yang tidak terlihat langsung. (Mor et al., 2021)

2.1.3.2 Set Keadaan Observasi (*Observable States*)

Kumpulan observasi yang dihasilkan oleh *hidden states*. Misalnya, dalam pengenalan ucapan, *observable states* dapat berupa rangkaian suara atau sinyal akustik. (Mor et al., 2021)

2.1.3.3 Matriks Transisi Keadaan (*Transitional Matrix*)

Matriks A yang mendefinisikan probabilitas transisi antar *hidden states*. Elemen a_{ij} dari matriks ini menunjukkan probabilitas berpindah dari *state i* ke *state j*. (Mor et al., 2021)

$$A = \{a_{ij}\}$$

Dimana $a_{ij} = P(q_{t+1} = j | q_t = i)$

Rumus 2. 2 Rumus Matriks Transisi Keadaan

Sumber: *ScienceDirect.com*

Keterangan.

A : matriks transisi (*transition matrix*) pada *Hidden Markov Model*.

a_{ij} : probabilitas berpindah dari *state i* pada waktu t ke *state j* pada waktu $t + 1$.

q_t : *state* pada waktu t .

2.1.3.4 Matriks Emisi (*Emission Matrix*)

Matriks B yang mendefinisikan probabilitas menghasilkan observasi tertentu dari *hidden states*. Elemen $b_j(O_t)$ menunjukkan probabilitas *state* j menghasilkan observasi O_t . (Mor et al., 2021)

$$B = \{b_j(k)\}$$

Dimana $b_j(k) = P(O_t = k | q_t = j)$

Rumus 2. 3 Rumus Matriks Emisi

Sumber: *ScienceDirect.com*

Keterangan.

B : matriks probabilitas observasi (*observation probability matrix*) pada *Hidden Markov Model*.

$b_j(k)$: probabilitas menghasilkan observasi k pada waktu t , jika sistem sedang berada pada *state* j .

O_t : observasi (output yang terlihat) pada waktu t .

q_t : *state* tersembunyi pada waktu t .

2.1.3.5 Distribusi Awal (*Initial Distribution*)

Vektor π yang mendefinisikan probabilitas awal berada di setiap *hidden states*. Elemen π_i menunjukkan probabilitas memulai di *state* i . (Mor et al., 2021)

$$\pi = \{\pi_i\}$$

Dimana $\pi_i = P(q_1 = i)$

Rumus 2. 4 Rumus Distribusi Awal

Sumber: *ScienceDirect.com*

Keterangan.

π : vektor probabilitas awal (*initial state distribution*) pada *Hidden Markov Model*.

π_i : probabilitas bahwa proses dimulai pada state i pada waktu $t = 1$.

q_1 : *state* tersembunyi pertama (awal).

2.1.4 Analisa *Data Time Series*

2.1.4.1 Komponen *Data Time Series*

Analisis *data time series* memerlukan identifikasi dan pemisahan dari berbagai komponen untuk mendapatkan pemahaman yang lebih baik tentang perilaku data dari waktu ke waktu (Chatfield, 2016). *Data time series* terdiri dari serangkaian observasi yang diambil secara berurutan dari waktu ke waktu. Ada beberapa komponen utama dalam *data time series* yang harus diperhatikan:

1. *Trend*

Komponen yang menunjukkan arah umum pergerakan data dalam jangka panjang. Misalnya, peningkatan harga rumah selama beberapa dekade.

2. Siklus (*Cycle*)

Pola yang terjadi pada interval waktu tertentu yang lebih panjang dari satu tahun, sering kali terkait dengan kondisi ekonomi atau bisnis. Misalnya, siklus ekonomi yang terjadi setiap 5-10 tahun.

3. Musiman (*Seasonal*)

Fluktuasi yang berulang secara reguler dalam interval waktu yang tetap, seperti bulanan atau tahunan. Misalnya, peningkatan penjualan es krim selama musim panas.

4. Irregular (*Residual / Noise*)

Komponen acak atau tidak teratur yang tidak dapat dijelaskan oleh *trend*, siklus, atau musiman. Ini biasanya dianggap sebagai noise atau variasi acak dalam data.

2.1.4.2 Metode Dekomposisi *Time Series*

Dekomposisi *time series* adalah teknik untuk memecah *data time series* menjadi komponen-komponen penyusunnya: *trend*, siklus, musiman, dan *irregular*. Ada beberapa metode untuk melakukan dekomposisi ini:

1. Dekomposisi Aditif (*Additive Decomposition*)

Asumsinya adalah bahwa *data time series* y_t dapat dinyatakan sebagai penjumlahan dari komponen-komponennya:

$$y_t = T_t + S_t + C_t + I_t$$

Rumus 2. 5 Rumus Dekomposisi Aditif

Sumber: *ScienceDirect.com*

Di mana T_t adalah trend, S_t adalah komponen musiman, C_t adalah siklus, dan I_t adalah *irregular*.

2. Dekomposisi Multiplikatif (*Multiplicative Decomposition*)

Asumsinya adalah bahwa *data time series* y_t dapat dinyatakan sebagai penjumlahan dari komponen-komponennya:

$$y_t = T_t \times S_t \times C_t \times I_t$$

Rumus 2. 6 Rumus Dekomposisi Multiplikatif

Sumber: *ScienceDirect.com*

Di mana T_t adalah trend, S_t adalah komponen musiman, C_t adalah siklus, dan I_t adalah *irregular*.

3. Metode STL (*Seasonal-Trend Decomposition using Loess*)

STL adalah metode yang sangat serbaguna dan dapat diterapkan pada berbagai jenis *data time series* dengan pola musiman dan trend yang kompleks (Chatfield, 2016). Metode non-parametrik yang fleksibel untuk mendekomposisi *data time series* menjadi komponen *trend*, musiman, dan residual menggunakan *local regression (Loess)*.

2.1.4.3 Model Prediksi *Time Series*

Pemilihan model yang tepat untuk prediksi *time series* bergantung pada karakteristik data dan tujuan prediksi, dengan *model ARIMA* dan metode eksponensial smoothing sering menjadi pilihan yang efektif untuk banyak aplikasi praktis (Hyndman & Athanasopoulos, n.d.). Ada beberapa model yang digunakan untuk prediksi *data time series*, yang dapat dikelompokkan menjadi dua kategori utama: model statistik dan *model machine learning*.

1. *Model ARIMA (Autoregressive Integrated Moving Average)*

Model yang menggabungkan *autoregressive (AR)*, *differencing (I)*, dan *moving average (MA)* untuk menangani data non-stasioner dan memprediksi nilai masa depan.

2. *Model Exponential Smoothing*

Menggunakan rata-rata tertimbang dari observasi sebelumnya dengan bobot yang menurun secara eksponensial. Contohnya adalah metode *Holt-Winters* yang memperhitungkan komponen *trend* dan musiman.

3. *Model GARCH (Generalized Autoregressive Conditional Heteroskedasticity)*

Digunakan untuk model volatilitas dalam *data time series* keuangan, dengan memprediksi varians data.

4. *Recurrent Neural Networks (RNN)*

Jaringan saraf yang dirancang untuk menangani data sekuensial, di mana output dari waktu sebelumnya digunakan sebagai input untuk waktu berikutnya.

5. *Long Short-Term Memory (LSTM)*

Jenis khusus dari *RNN* yang mampu menangkap hubungan jangka panjang dalam data *time series* dengan mengatasi masalah *vanishing gradient*.

6. *Random Forest dan Gradient Boosting*

Algoritma *ensemble* yang dapat digunakan untuk prediksi *time series* dengan menangani fitur waktu secara eksplisit.

2.1.5 **Prediksi Harga *Cryptocurrency***



Gambar 2. 2 *Bitcoin*

Sumber: *Google.com*

Prediksi harga *cryptocurrency* merupakan tantangan signifikan dalam bidang keuangan dan teknologi informasi, mengingat volatilitas tinggi dan kompleksitas pasar aset digital. Penelitian terkini menunjukkan bahwa pendekatan berbasis *machine learning* dan *deep learning* mampu memberikan hasil yang lebih akurat dibandingkan metode statistik tradisional.

Dalam studi oleh (Bouteska et al., 2024), dilakukan analisis komparatif antara *model ensemble learning* dan *deep learning* untuk peramalan harga berbagai *cryptocurrency*. Hasilnya menunjukkan bahwa *model ensemble learning*, seperti *Random Forest* dan *Gradient Boosting*, serta *model deep learning*, seperti *LSTM* dan *GRU*, memiliki kinerja yang unggul dalam memprediksi harga *cryptocurrency*.

Selain itu, penelitian oleh (Zhao, 2020) mengembangkan kerangka kerja *deep learning* berbasis *Long Short-Term Memory (LSTM)* untuk memprediksi pergerakan harga aset digital menggunakan data perdagangan per transaksi. Model ini menunjukkan akurasi tinggi dalam memprediksi perubahan harga jangka pendek, dengan kinerja yang stabil pada periode pengujian di luar sampel.

Pendekatan lain yang menarik adalah penggunaan sinyal sosial untuk meningkatkan akurasi prediksi. (Glenski et al., 2019) memanfaatkan data dari platform sosial seperti *Reddit* dan *GitHub* untuk memprediksi harga *Bitcoin*, *Ethereum*, dan *Monero*. Model yang menggabungkan data harga historis dan sinyal sosial menunjukkan peningkatan akurasi dibandingkan model yang hanya menggunakan data harga.

Dalam konteks algoritma, studi oleh (Biswas et al., 2021) membandingkan kinerja *LSTM*, *GRU*, dan *Bi-LSTM* dalam pendekatan multivariat untuk prediksi harga *cryptocurrency*. Hasilnya menunjukkan bahwa *Bi-LSTM* dan *GRU* memiliki kinerja yang lebih baik dibandingkan *LSTM* dalam memodelkan data time series harga *cryptocurrency*.

Secara keseluruhan, penelitian-penelitian tersebut menekankan pentingnya pemilihan model dan fitur yang tepat dalam prediksi harga *cryptocurrency*. Integrasi data historis, sinyal sosial, dan penggunaan algoritma *machine learning* serta *deep learning* yang sesuai dapat meningkatkan akurasi prediksi, yang sangat bermanfaat bagi investor dan pelaku pasar dalam pengambilan keputusan.

2.1.6 Penerapan *Hidden Markov Model* dalam Analisis Keuangan

2.1.6.1 Penerapan *HMM* untuk Analisis Pasar Keuangan

Hidden Markov Model (HMM) digunakan dalam analisis pasar keuangan untuk memodelkan dan memprediksi dinamika harga aset. *HMM* memungkinkan para analis untuk menangkap pola tersembunyi dalam data keuangan, yang dapat berupa perubahan dalam keadaan pasar seperti *bull markets* (pasar naik) dan *bear markets* (pasar turun).

Penerapan *HMM* dalam keuangan melibatkan identifikasi keadaan tersembunyi yang menggambarkan kondisi pasar yang berbeda dan penggunaan transisi antar keadaan ini untuk memprediksi harga masa depan. *HMM* dapat menggabungkan informasi dari berbagai indikator teknis dan fundamental untuk memberikan prediksi yang lebih akurat.

Hidden Markov Model digunakan secara luas di pasar keuangan karena kemampuannya untuk memodelkan data *time series* yang menunjukkan dinamika yang berubah seiring waktu (Bosire & Maina, 2021). Model keuangan tradisional sering kali mengasumsikan stasioneritas, yang mungkin tidak secara akurat mencerminkan kondisi pasar yang sebenarnya. Di sisi lain, *HMM* dapat beradaptasi dengan lingkungan yang tidak stasioner dengan memodelkan keadaan yang dapat diamati yang dipengaruhi oleh rezim tersembunyi (Oelschläger & Adam, 2021).

2.1.6.2 Studi Kasus Penggunaan *HMM* dalam Prediksi Harga Saham dan Aset Keuangan Lainnya

2.1.6.2.1 Prediksi Harga Saham

Salah satu studi kasus terkenal adalah penggunaan *HMM* untuk prediksi harga saham. Dalam penelitian ini, harga saham dikategorikan ke dalam beberapa keadaan tersembunyi yang mewakili berbagai fase pasar. Setiap keadaan memiliki distribusi probabilitas yang berbeda untuk harga saham, yang kemudian digunakan untuk memprediksi pergerakan harga di masa depan.

Misalnya, studi oleh (Nguyen & Nguyen, 2015) menunjukkan bahwa "*HMM can be effectively used to model and predict stock prices, where different hidden states correspond to bullish, bearish, and stagnant market conditions*". Penelitian ini berhasil menunjukkan bahwa *HMM* dapat memberikan prediksi yang lebih baik dibandingkan dengan model tradisional seperti ARIMA dalam beberapa kasus.

2.1.6.2.2 Prediksi Harga Aset Lainnya

Selain saham, *HMM* juga digunakan untuk memprediksi harga aset keuangan lainnya seperti komoditas, mata uang, dan *cryptocurrency*. Misalnya, dalam analisis harga *Bitcoin*, *HMM* dapat digunakan untuk mengidentifikasi keadaan pasar yang berbeda seperti periode volatilitas tinggi dan rendah, yang kemudian digunakan untuk memprediksi pergerakan harga di masa depan.

Penelitian oleh Shen dan Shafiq (2020) menunjukkan bahwa "*HMM can effectively capture the regime-switching behavior of Bitcoin prices, providing valuable insights for traders and investors*". Penelitian ini mengindikasikan bahwa *HMM* dapat membantu dalam memahami dan memprediksi dinamika harga *Bitcoin* yang kompleks.

2.1.6.3 Keuntungan dan Keterbatasan *HMM* dalam Konteks Keuangan

2.1.6.3.1 Keuntungan *HMM*

1. Modelling Hidden states

HMM memungkinkan identifikasi dan pemodelan keadaan tersembunyi dalam pasar keuangan, yang dapat memberikan wawasan lebih dalam tentang dinamika pasar.

2. Flexibilitas

HMM dapat diterapkan pada berbagai jenis data keuangan dan dapat disesuaikan untuk menangkap pola kompleks dalam data time series.

3. Probabilistic Framework

HMM menyediakan kerangka probabilistik untuk prediksi, yang memungkinkan pengukuran ketidakpastian dan risiko yang lebih baik.

2.1.6.3.2 Keterbatasan *HMM*

1. Kompleksitas Komputasi
2. Pelatihan *HMM* dapat menjadi sangat kompleks dan memerlukan sumber daya komputasi yang signifikan, terutama untuk dataset yang besar.
3. Sensitivitas terhadap Parameter

Hasil *HMM* sangat bergantung pada pemilihan parameter awal dan struktur model, yang dapat mempengaruhi akurasi prediksi.

4. Overfitting

Salah satu penyebab mendasar *overfitting* dalam *HMM* adalah estimasi status tersembunyi yang tidak tepat. Ketika model menyertakan sejumlah status yang berlebihan tanpa dukungan data yang memadai, model tersebut cenderung mengalami *overfitting*, karena model tersebut berupaya menjelaskan variabilitas dalam set data pelatihan alih-alih berfokus pada pola yang dapat digeneralisasi (van Havre et al., 2016). Fenomena ini khususnya terlihat dalam sistem yang menggunakan model campuran *Gaussian* untuk merepresentasikan probabilitas emisi, yang dapat menciptakan model kompleks yang tidak dapat digeneralisasi dengan baik (Raghavan et al.,)

2.1.7 Teknik Pra-Pemrosesan Data

Pra-pemrosesan data adalah aspek fundamental sebelum penerapan model *Hidden Markov Model (HMM)*. Tujuannya adalah memastikan kualitas data sebaik mungkin (bebas *noise*, *outlier*, atau nilai hilang) agar model bisa menangkap pola pasar secara optimal.

2.1.7.1 Penanganan *Missing Value*

Nilai hilang dapat menyebabkan bias dan melemahkan kinerja estimasi model *HMM*. Strategi yang umum digunakan:

1. *Drop row* saat jumlah missing kecil, atau
2. Imputasi dengan metode seperti *last observation carried forward (LOCF)*, interpolasi *linear* atau *spline*, atau *multiple imputation*.

Artikel oleh (Kirilin, 2024) menekankan bahwa *LOCF* sering efektif pada data keuangan namun perlu hati-hati terhadap potensi artefak *noise*. Sedangkan studi (BMC Medical, 2024) mengingatkan bahwa penghapusan baris secara langsung dapat membuat *data time series* menjadi tak seimbang dan melanggar asumsi kontinuitas waktu.

2.1.7.2 Penghapusan *Noise* dan *Outlier*

Outlier harga ekstrim (seperti *flash crash* atau data salah) dapat merusak estimasi parameter *Gaussian* pada *HMM*. Teknik smoothing yang umum:

1. *Moving Average (MA/SMA/EMA)*
2. *Wavelet* atau *decomposition* seperti *EMD*

Menurut (Investopedia, 2023), *moving average* berfungsi sebagai *low-pass filter* untuk menghaluskan *data noise* tanpa menghilangkan tren penting. Dalam trading kripto, *moving averages* membantu mengekstrak pola jangka panjang dari data yang sangat *volatile*.

2.1.7.3 Transformasi Stasionaritas dan Normalisasi

Model HMM umumnya mengasumsikan distribusi data yang stasioner. Teknik yang digunakan pada umumnya seperti *differencing* untuk membuat mean menjadi tetap, standarisasi (*Z-Score*), dan *min-max scaling*. Transformasi ini mengurangi *skewness* dan variansi ekstrem, sehingga model lebih stabil di seluruh dataset.

2.1.7.4 Resampling dan Agregasi Data

Frekuensi data (1-menit, jam, harian) harus disesuaikan dengan tujuan model. Data 1-menit cocok untuk *microstructure*, namun *noise* tinggi. Data harian hingga mingguan lebih cocok untuk mendeteksi regime pasar jangka panjang. Pemilihan frekuensi ini mempengaruhi jumlah data (*sample size*) serta tingkat kebisingan (*noise*) yang ingin direspon model.

2.1.7.5 Ekstraksi Fitur Teknis

Dengan data yang sudah bersih, fitur teknikal ditambahkan untuk mendukung pemisahan *state* pasar. *Moving Average (MA)* untuk identifikasi tren. *Volume* untuk indikasi minat pasar. Return harian untuk ukuran momentum. Tambahan opsional seperti *RSI*, *MACD*, dan

lain-lain. Penyertaan indikator ini terbukti meningkatkan akurasi model di penelitian pasar keuangan.

2.1.7.6 Normalisasi Fitur

Karena rentang variabel berbeda, normalisasi sangat penting. Jika tidak, fitur volume yang besar dapat mendominasi model dan standarisasi memastikan model memproses setiap fitur secara seimbang.

2.1.7.7 Implementasi Pra-Pemroses dalam Kode

```
python
# 1. Timestamp dan filtering
data['Timestamp'] = pd.to_datetime(data['Timestamp'], unit='s')
data = data[(data['Timestamp'] >= '2020-01-01') & (data['Timestamp'] < '2025-01-01')]

# 2. Moving average smoothing & dropna
data['MA'] = data['Close'].rolling(window=200).mean()
data.dropna(inplace=True)

# 3. Standarisasi fitur
from sklearn.preprocessing import StandardScaler
features = StandardScaler().fit_transform(data[['Close', 'MA', 'Volume']])
```

Gambar 2. 3 Kode Pra-Pemroses

Sumber: Data Peneliti, 2025

Langkah ini menjamin bahwa data model stabil (tidak ada missing), tidak *noisy*, dan disebar secara seimbang untuk pelatihan *HMM*.

2.2 Penelitian Terdahulu

Pada bagian ini, saya akan mengulas beberapa penelitian terdahulu yang relevan dengan topik penelitian tentang penggunaan *Hidden Markov Model (HMM)* untuk prediksi harga *Bitcoin*. Penelitian-penelitian yang diulas memiliki kaitan

dengan judul, metode, dan masalah penelitian yang diambil, serta memuat informasi mengenai permasalahan, indikator, metode, tools, output, kendala, dan hal-hal lain yang relevan.

N o.	Judul Penelitian	Penulis & Tahun	Permasalahan	Metode & Algoritma	Output	Kendala & Catatan
1.	Prediksi Harga Cryptocurrency Menggunakan Algoritma Long Short Term Memory (LSTM)	Moch Farryz Rizkilloh, Sri Widiyanti, 2022	Prediksi harga cryptocurrency, khususnya <i>Bitcoin</i> , yang sangat fluktuatif dan spekulatif	Long Short Term Memory (LSTM)	Model prediksi harga cryptocurrency dengan Tingkat akurasi yang baik, diukur menggunakan RMSE	Tantangan dalam mengelola data time series jangka panjang dan mengoptimalkan parameter model
2.	Penerapan <i>Hidden Markov Model</i> pada Harga Saham	Sri Wahyuni Mamonto, Yohanes A. R. Langi, Altien J. Rindegan, 2016	Prediksi peluang kenaikan harga saham dari beberapa bank besar di Indonesia menggunakan data harga saham harian	<i>Hidden Markov Model (HMM)</i> , Algoritma Baum-Welch, Algoritma Viterbi	Prediksi peluang kenaikan harga saham untuk periode satu minggu, satu bulan, dan satu tahun.	<i>HMM</i> memprediksi peluang kenaikan saham tertinggi mingguan dan bulanan ada pada Bank Mandiri, sedangkan tahunan 2016 pada Bank BRI. Algoritma Viterbi menunjukkan tren penurunan, namun saham BRI tetap

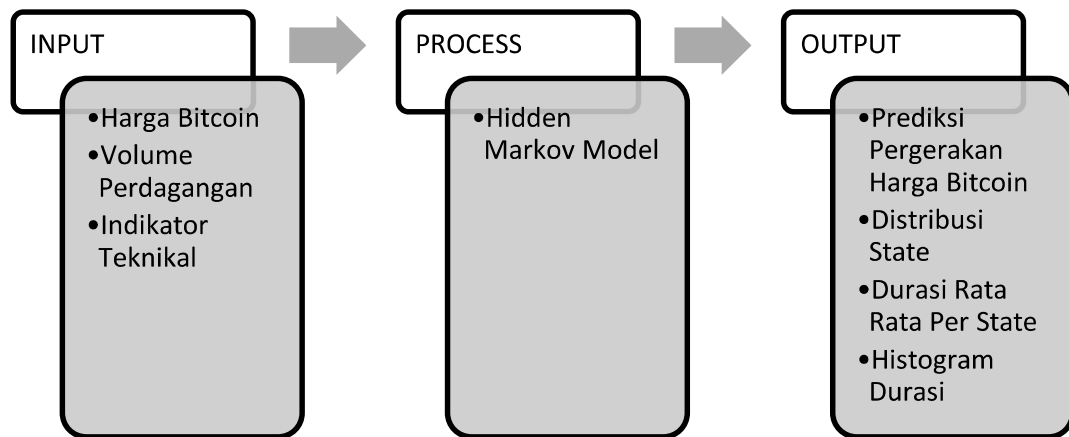
						berpeluang naik.
3.	Prediction of <i>Bitcoin</i> Price Based on the <i>Hidden Markov Model</i>	Muxi Li, 2021	Prediksi harga <i>Bitcoin</i> menggunakan <i>HMM</i> dengan data harga penutupan dan volume perdagangan	<i>Hidden Markov Model (HMM)</i>	Prediksi harga <i>Bitcoin</i> untuk jangka pendek, menengah, dan Panjang	Akurasi prediksi menurun seiring dengan panjang periode prediksi
4.	<i>Bitcoin</i> Price Forecast Using LSTM and GRU Recurrent Networks, and <i>Hidden Markov Model</i>	Yike Xu, 2020	Perbandingan performa <i>HMM</i> dengan LSTM dan GRU dalam memprediksi harga <i>Bitcoin</i>	<i>HMM</i> , LSTM, GRU	Evaluasi performa model menggunakan MAPE dan RMSE	<i>HMM</i> dengan Gaussian Mixture Model menunjukkan performa terbaik
5.	Predicting <i>Bitcoin</i> Prices Using <i>Hidden Markov Model</i>	Kagame Richard, Aime Munezero, Samiratu Ntohs Daniel Emesiani, 2022	Optimalisasi jumlah <i>hidden states</i> dalam <i>HMM</i> untuk prediksi harga <i>Bitcoin</i>	<i>Hidden Markov Model (HMM)</i> , Bayesian Information Criterion (BIXC)	Prediksi harga <i>Bitcoin</i> dengan jumlah <i>hidden states</i> optimal	<i>HMM</i> menunjukkan performa lebih baik dibandingkan dengan Artificial Neural Networks
6.	Expectile <i>Hidden Markov Model</i> Regression Models for Analyzing Cryptocurrency Returns	Beatrice Foroni, Luca Merlo, Lea Petrella, 2023	Pengembangan model regresi <i>HMM</i> untuk analisis return cryptocurrency	Expectile Hidden Markov Regression Models, Expectation – Maximization	Analisis return harian <i>Bitcoin</i> dan hubungannya dengan indeks	Fokus pada return ekstrem dan evolusinya seiring waktu

			dalam kerangka manajemen risiko	Algoritma	pasar dunia	
7.	<i>Bitcoin Hidden Markov Model Analysis</i>	Andrei Crapotca, 2024	Analisis pergerakan harga <i>Bitcoin</i> menggunakan Variational Gaussian <i>HMM</i>	Variational Gaussian <i>Hidden Markov Model</i> , <i>HMMlearn</i> Python Library	Analisis pergerakan harga <i>Bitcoin</i> dari 2017 hingga 2023	Model dapat menangani data kontinu yang kompleks dan menganalisis beberapa aliran data secara bersamaan
8.	Decoding Crypto Volatility: Leveraging <i>Hidden Markov Models</i> for Predictive Insights	Christina Chiaro, 2024	Pemodelan volatilitas pasar cryptocurrency menggunakan <i>HMM</i> untuk prediksi pergerakan harga	<i>Hidden Markov Model (HMM)</i>	Prediksi pergerakan harga cryptocurrency berdasarkan volatilitas pasar	Memberikan wawasan tentang dinamika pasar cryptocurrency yang kompleks
9.	Exploring the Predictability of Cryptocurrency via Bayesian <i>Hidden Markov Model</i>	Constantina Koki, Stefanos Leonardos, Georgios Piliouras, 2020	Analisis prediktabilitas harga cryptocurrency menggunakan Bayesian <i>Hidden Markov Model</i>	Bayesian <i>Hidden Markov Models</i> , <i>Multi-state HMM</i>	Evaluasi prediksi harga <i>Bitcoin</i> , <i>Ether</i> , dan <i>Ripple</i> dengan <i>HMM</i>	Mengidentifikasi periode pasar dengan karakteristik return yang berbeda

Tabel 2. 1 Penelitian Terdahulu

Sumber: Data Peneliti, 2025

2.3 Kerangka Pemikiran



Gambar 2. 4 Kerangka Pemikiran

Sumber: Data Peneliti, 2025

Kerangka pemikiran ini bertujuan untuk memberikan gambaran tentang bagaimana *Hidden Markov Models (HMM)* dapat digunakan untuk memprediksi harga *Bitcoin* dengan mempertimbangkan faktor-faktor seperti harga perdagangan, volume perdagangan dan indikator teknis. Kerangka ini akan menjelaskan hubungan antara variabel-variabel utama, serta bagaimana penerapan *HMM* dapat memberikan solusi untuk memecahkan masalah yang telah diidentifikasi.

Penelitian ini dimulai dengan mengumpulkan data berupa volume perdagangan, harga historis *Bitcoin*, dan indikator teknis sebagai komponen utama. Seluruh data ini diproses menjadi input bagi algoritma *Hidden Markov Model (HMM)*. Model *HMM* digunakan untuk mengidentifikasi pola tersembunyi dalam data, yang kemudian menghasilkan prediksi harga *Bitcoin*. Hasil prediksi tersebut

selanjutnya dianalisis lagi yang akan menghasilkan distribusi *state*, durasi rata-rata per *state*, dan histogram durasi.