

BAB III

METODOLOGI PENELITIAN

3.1 Gambaran Umum Objek Penelitian

Dalam penelitian, objek yang akan diteliti adalah 3 jenis *cryptocurrency* dengan peringkat kapitalisasi pasar yang mendominasi di tahun 2024. Peringkat kapitalisasi pasar diperoleh saat awal penelitian berlangsung pada tanggal 7 Februari 2024 dari *website* resmi coinmarketcap.com, yang merupakan *website* dengan sumber informasi terpercaya, komprehensif, dan kerap digunakan di dalam dunia *cryptocurrency*.

Data historis berupa harga penutupan (*closing price*) dari ketiga *cryptocurrency* digunakan sebagai dasar dalam pembangunan model prediksi. Data diperoleh dari Yahoo Finance, *website* bagian resmi dari Yahoo yang digunakan sebagai sumber informasi dan data berita keuangan, termasuk saham, *cryptocurrency*, dan bahkan keuangan pribadi. Pemilihan *website* Yahoo Finance sebagai sumber memperoleh data historis *cryptocurrency* didasarkan oleh kelengkapannya dan popularitasnya yang tinggi dikalangan para analis dan penelitian-penelitian sebelumnya. Adapun dukungan teknis dari penggunaan *website* Yahoo Finance terhadap proses pembangunan model, yaitu *library yfinance* dan dukungan *library* terhadap bahasa pemrograman Python. Dengan adanya *library* khusus ini, peneliti dapat secara langsung mengakses data dari Yahoo Finance dan mengintegrasikan data dalam *script* Python.

Dalam penelitian ini terdapat dua pilihan model *machine learning*, yaitu *Extreme Gradient Boosting* (XGBoost) dan *Light Gradient Boosting Machine* (LightGBM), serta dua pilihan model *deep learning*, yaitu *Convolutional Neural Networks* (CNN) dan *Long Short-Term Memory* (LSTM). Setiap model akan dikembangkan dan diuji performanya sesuai dengan tahapan penelitian yang sudah ditetapkan untuk menentukan efektivitasnya dalam menganalisa dan memprediksi pergerakan harga *cryptocurrency*.

Pada tahapan awal penelitian, kedua algoritma *deep learning* akan dikembangkan untuk dijadikan sebagai model untuk memprediksi harga *cryptocurrency* secara terpisah. Evaluasi dan perbandingan akan dilakukan untuk melihat efektivitas, keoptimalan dan secara keseluruhan menentukan algoritma *deep learning* yang memiliki performa lebih unggul. Selanjutnya, algoritma dengan hasil terbaik akan diintegrasikan dengan masing-masing pilihan model *machine learning*, yaitu *Extreme Gradient Boosting* (XGBoost) dan *Light Gradient Boosting Machine* (LightGBM) untuk membentuk model *hybrid*.

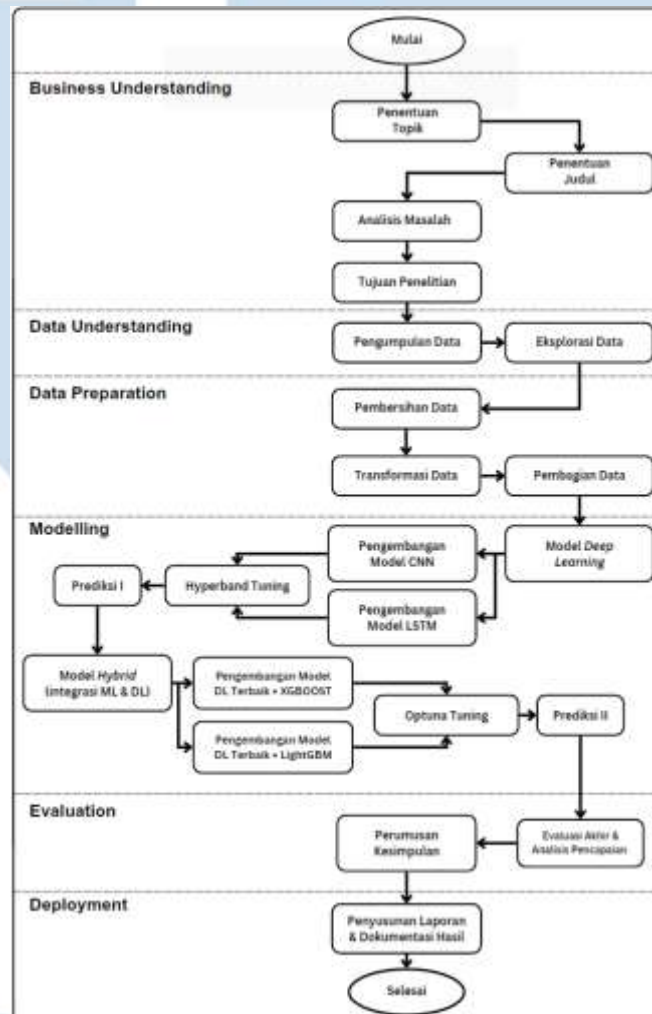
Kedua konfigurasi model *hybrid* yang mengkombinasikan (1) algoritma *deep learning* terbaik dengan *Extreme Gradient Boosting* (XGBoost) dan (2) kombinasi algoritma *deep learning* terbaik dengan *Light Gradient Boosting Machine* (LightGBM), akan melalui proses tuning yang bertujuan untuk menghasilkan sebuah model prediktif yang memiliki tingkat keoptimalan dan akurasi yang. Selanjutnya terdapat tahapan terakhir yang sangat krusial dalam penelitian, yaitu tahap evaluasi dan perbandingan antara kedua model *hybrid*. Tujuan dari tahapan ini adalah untuk menentukan model *hybrid* yang memiliki performa terbaik, serta membuktikan kelebihan signifikan dibandingkan dengan penggunaan model tunggal.

3.2 Metode Penelitian

Penelitian yang dilakukan dapat digolongkan ke dalam penelitian metode kuantitatif, yang merupakan penelitian berfokus pada pengumpulan dan analisa data yang berbentuk numerik [58]. Data yang digunakan adalah data yang berbentuk numerik, termasuk tanggal dan nilai-nilai harga *cryptocurrency* (numerik). Evaluasi hasil performa dari model yang dikembangkan juga didasarkan dengan penggunaan rumus matematika untuk metrik evaluasi. Dengan demikian seluruh proses penelitian secara intensif melibatkan analisis kuantitatif dengan berfokus pada analisa data numerik.

3.2.1 Alur Penelitian

Pada gambar 3.1 terdapat alur penelitian yang telah dibuat dalam bentuk *flowchart diagram*. Dengan adanya gambaran dari alur penelitian, diharapkan pemahaman mengenai alur (*end-to-end*) penelitian dapat dimengertikan secara jelas.

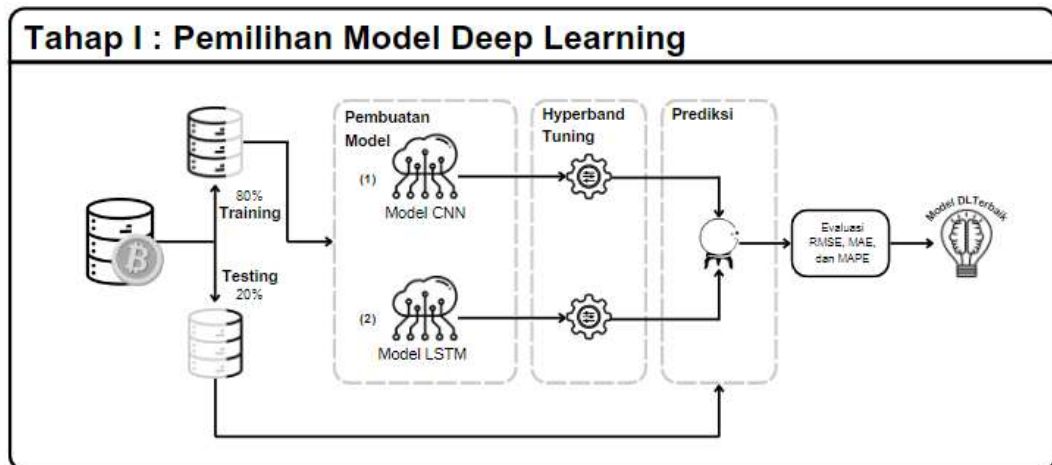


Gambar 3. 1 Alur Penelitian CRISP-DM

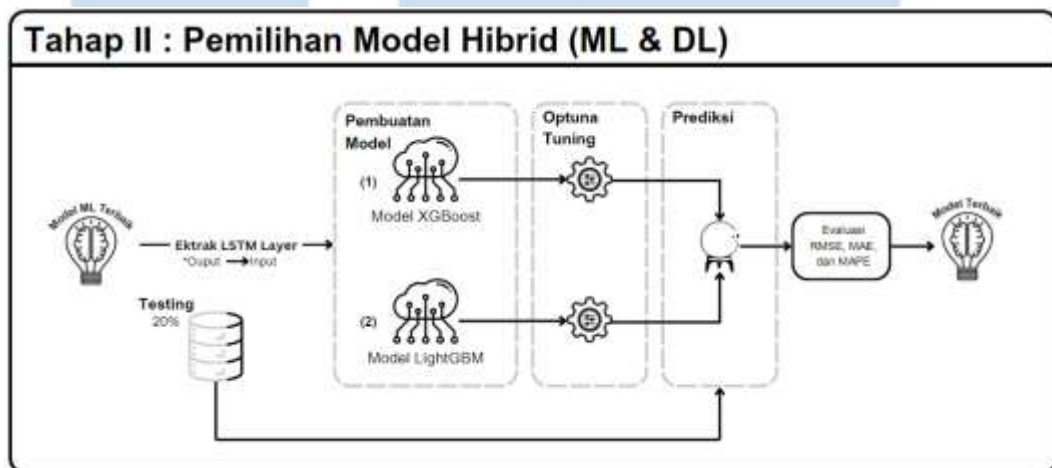
Alur penelitian diilustrasikan dalam Gambar 3.1, dimulai dengan proses *business understanding* yang menjadi landasan terhadap target dan solusi dalam menghadapi pergerakan harga *cryptocurrency* yang sangat fluktuatif. Pada tahapan ini diperlukan pemahaman mendalam mengenai prediksi fluktuasi harga *cryptocurrency*. Setelah menetapkan dasar penelitian, langkah yang akan dilakukan berikutnya adalah *data understanding*. Tahapan ini meliputi

pengumpulan data historis dan analisis awal untuk mendeteksi pola-pola penting yang mempengaruhi prediksi harga *cryptocurrency*. Kemudian, terdapat fase yang krusial dalam penelitian, yaitu *data preparation*. Pada tahapan *data preparation*, data akan dibagi dan disesuaikan agar menghasilkan data yang terstruktur dan bersih untuk diprediksi.

Tahapan *modelling* dalam penelitian dapat dilihat secara lebih rinci pada Gambar 3.2 dan 3.3. Dalam Gambar 3.2, kedua model *deep learning*, yaitu CNN dan LSTM akan dikembangkan dan dievaluasi untuk menentukan model terbaik dengan mengimplementasikan teknik Hyperband *tuning*. Hasil evaluasi akan ditentukan dari RMSE, MAE, dan MAPE. Model terbaik akan digunakan dengan diintegrasikan ke dalam kedua konfigurasi model *machine learning*, XGBoost dan LightGBM, seperti yang terdapat pada Gambar 3.3. Kedua model akan melalui proses *tuning* kembali dengan menggunakan teknik yang berbeda, yaitu teknik Optuna. Proses *evaluation* akan kembali dilakukan untuk menentukan model *hybrid* dengan tingkat keoptimalan dan keakuratan yang terbaik. Pada tahapan ini, analisa dan perumusan kesimpulan juga dilakukan, karena akan ditemukan pada tahap terakhir, yaitu deployment. Dalam tahapan terakhir, hasil yang akan diperoleh setelah penelitian telah dijalankan adalah laporan dan dokumentasi yang akan dijadikan sebagai sumber informasi untuk analisis lebih mendalam ataupun praktisi secara langsung dalam dunia *cryptocurrency*. Skripsi ini diharapkan dapat memperlihatkan potensi akan penggunaan model *hybrid* dalam dunia *cryptocurrency* kedepannya, serta menjadi wawasan tambahan untuk penelitian masa depan dalam bidang *machine learning* dan *deep learning*.



Gambar 3. 2 Kerangka Tahap I Proses Pemilihan Model Deep Learning



Gambar 3. 3 Kerangka Tahap II Proses Pemilihan Model *Hybrid* (ML & DL)

3.2.2 Metode Data Mining

Terdapat 2 dua jenis metodologi yang akan diimplementasikan dalam penelitian yaitu *Data Mining Methodology* dan *Problem Solving Methodology*. Berikut merupakan tabel yang membandingkan *Data Mining Methodology*, yaitu SEMA, dan CRISP-DM:

UNIVERSITAS
MULTIMEDIA
NUSANTARA

Tabel 3. 1 Perbandingan Data Mining Methodology

Indikator	SEMMA [59], [60]	CRISP-DM [41]
Langkah-langkah	<ol style="list-style-type: none"> 1. <i>Sample</i> 2. <i>Explore</i> 3. <i>Modify</i> 4. <i>Model</i> 5. <i>Assessment</i> 	<ol style="list-style-type: none"> 1. <i>Business Understanding</i> 2. <i>Data Understanding</i> 3. <i>Data Preparation</i> 4. <i>Modeling</i> 5. <i>Evaluation</i> 6. <i>Deployment</i>
Kelebihan	<ul style="list-style-type: none"> • Alat fungsional dari <i>SAS Enterprise Miner</i> • Terpusat dengan <i>SAS Enterprise Miner</i> • Dapat digunakan pada berbagai jenis proyek • Memanfaatkan berbagai teknik statistic untuk analisis mendalam 	<ul style="list-style-type: none"> • Mudah untuk diimplementasikan • Dapat memahami data secara lebih unggul • Bersifat umum (<i>general</i>) • Fleksibel karena dapat digunakan pada variasi data yang luas
Kekurangan	<ul style="list-style-type: none"> • Tidak adanya proses pemahaman data (minim analisa) • Proses <i>Deployment</i> tidak dilakukan • Mwmbutuhkan waktu dan sumber daya yang signifikan • Tidak ideal untuk data yang terus berubah 	<ul style="list-style-type: none"> • Terdiri dari berbagai tahapan • Sangat iterative (bergantungan pada masalah), maka dapat memakan waktu

Berdasarkan tabel perbandingan berikut, dapat disimpulkan bahwa Data Mining Methodology yang lebih unggul dan tepat untuk digunakan dalam penelitian ini adalah Cross-Industry Standard Process (CRISP-DM), walaupun kedua metodologi ini hampir serupa satu sama lain dan diturunkan dari KDD [61]. Pilihan juga didasari dari penelitian-penelitian sebelumnya yang menggunakan CRISP-DM sebagai pilihan *Data Mining Methodology*, seperti penelitian dari Derbentsev et al. dan Millikan et al., yang menekankan bahwa CRISP-DM merupakan proses standar yang secdara luas digunakan dan secara lengkap menggambarkan fase-fase dan metode-metode utama dalam pengembangan data [62], [63]. Berikut merupakan tahapan CRISP-DM yang diimplementasikan kedalam penelitian ini:

1. *Business Understanding*: Tahapan awal yang mengharuskan peneliti untuk untuk secara rinci mengobservasi situasi bisnis yang perlu dilihat

untuk memahami sumber daya tersedia dan sumber daya apa yang dibutuhkan. Salah satu hal terpenting yang harus dilakukan dalam fase ini adalah memutuskan apa tujuan dan target dari penelitian. Pertama, jenis data mining (seperti regresi) dan kriteria keberhasilan data mining harus dijelaskan (seperti akurasi prediksi). Tidak melupakan, bahwa tahapan ini diperlukan pemahaman yang mendasari pembuatan proyek, serta mengetahui masalah yang terdapat di dunia *cryptocurrency*. Dengan memahami asal-usul dari masalah, langkah selanjutnya dalam alur penelitian akan lebih mudah untuk dipahami.

2. *Data Understanding*: Fase yang dimana fokus utama adalah untuk mengumpulkan data dari sumber data terpercaya, mengeksplorasi dan mendeskripsikannya, serta memeriksa kualitas data. Untuk memberikan pemahaman yang mendalam dan detail mengenai data, visualisasi data akan diperlukan. Proses ini melibatkan penggunaan analisis statistik dan pengenalan terhadap atribut dan bagaimana data dapat dikelompokkan. Data yang dikelompokkan pada penelitian ini diperoleh dengan menggunakan *library yfinance* yang terdapat pada *Jupyter Notebook*. Pemahaman mengenai data sangat krusial sebelum penelitian berlangsung, ini esensial agar dapat mengetahui fungsi dan keperluan data, serta variabel mana yang akan digunakan dalam penelitian.
3. *Data Preparation*: Mencakup semua aktivitas yang diperlukan untuk menyusun kumpulan data akhir (data yang akan dimasukkan ke dalam alat pemodelan) dari data mentah awal. Ada beberapa hal yang akan dilakukan mencakup pembersihan data (*Data Cleaning*), melakukan pemilihan data (*Data Selection*), serta adanya transformasi terhadap data (*Data Transformation*) untuk dijadikan sebagai *input* dalam tahap pemodelan. Pada tahapan ini juga diperlukan proses dalam mengkonfirmasi data-data yang memerlukan perbaikan dengan menghilangkan (*null*) dan *errors* pada data.

4. *Modelling*: Dalam fase pemodelan data, metode pemodelan dipilih untuk kemudian dibangun kasus uji dan model. Secara umum, pilihan tergantung pada data dan masalah dalam penelitian. Cara terbaik untuk mengetahui seberapa bagus suatu model adalah membandingkannya dengan kriteria evaluasi dan memilih yang terbaik. Pilih dan terapkan berbagai teknik pemodelan, dan *tuning* parameter akan dilakukan agar mendapatkan model yang optimal dan akurat (dengan acuan kriteria evaluasi pilihan). Dalam penelitian kedua algoritma *deep learning*, yaitu CNN dan LSTM akan melalui proses *tuning* dengan menggunakan teknik *hyperband* untuk mendapatkan model yang optimal dan tepat untuk menangani data. Kedua model akan dievaluasi berdasarkan metrik performa yang telah ditetapkan yaitu RMSE, MAE dan MAPE. Model *deep learning* yang unggul akan diintegrasikan dengan metode *machine learning*, yaitu XGBoost dan LightGBM untuk menghasilkan model *hybrid*. Nantinya model *hybrid* akan dituning dengan teknik Optuna dan dievaluasi kembali (evaluasi akhir) dalam tahap *evaluation* untuk mendapatkan model *hybrid* yang optimal, akurat dan dapat mendukung keterbatasan model tunggal (secara terpisah).
5. *Evaluation*: Pada fase ini, hasil temuan dari tahapan *modelling* akan dikembangkan dan disesuaikan kembali dengan tujuan penelitian yang telah ditetapkan. Evaluasi model secara menyeluruh, dan peninjauan langkah-langkah yang dijalankan untuk menyusun model juga harus dilakukan untuk memastikan model tersebut mencapai tujuan dan standar dari penelitian. Terakhir, yaitu perumusan kesimpulan yang dikembangkan setelah mendapatkan hasil evaluasi menyeluruh terhadap keseluruhan proses penelitian. Perumusan kesimpulan memberikan pemahaman mendalam mengenai efektivitas dan keakuratan model.
6. *Deployment*: Fase terakhir dari CRISP-DM dan dapat berupa sesuatu yang sederhana seperti laporan akhir, perangkat lunak, dan sebagainya. Hasil akhir (*output*) dari penelitian berupa laporan, serta dokumentasi

model dan visualisasi hasil yang diharapkan dapat dijadikan referensi untuk penelitian selanjutnya ataupun implementasi praktis penggunaan model yang telah dirancang dalam penelitian.

3.2.3 Metode *Problem Solving*

3.2.3.1 Pilihan *Deep Learning* Model

Tabel 3. 2 Perbandingan Algoritma LSTM , GRU dan RNN

Indikator	Long Short-Term Memory Network (LSTM) [64]	Gated Recurrent Unit (GRU) [44], [65], [66]	Recurrent Neural Network (RNN) [67]
Kelebihan	<ul style="list-style-type: none"> • Memberikan gambaran yang memiliki urutan (kronologis) • Menyimpan informasi dalam jangka waktu lama • Lebih akurat saat menggunakan <i>dataset</i> yang memiliki urutan panjang 	<ul style="list-style-type: none"> • Kecepatan yang tinggi dan melebihi LSTM • Menggunakan memory yang minim 	<ul style="list-style-type: none"> • Cocok untuk digunakan dalam memproses data urutan • Dapat memberikan informasi jangka pendek dalam memori • Dapat menangani length dari input dan output
Kekurangan	<ul style="list-style-type: none"> • Memerlukan memori yang lebih • Terkadang <i>overfitting</i> menjadi problem utama • Variasi parameter yang luas yang harus dipelajari 	<ul style="list-style-type: none"> • Efisiensi pembelajaran yang rendah • Konvergensi lambat 	<ul style="list-style-type: none"> • Sulit untuk dipelajari dan menggabungkan data yang panjang. • Sering terdapat error pada gradien • Memerlukan suberdaya yang memadai • Sulit untuk dilakukannya <i>training</i>

Dalam tahapan awal penelitian ini, terdapat dua (2) model yang akan dikembangkan menggunakan algoritma *deep learning* yang dipilih berdasarkan keunikan dan keunggulan yang ditawarkan, serta kesesuaiannya dengan dataset *time series* yang kiranya akan digunakan. Algoritma pilihan pertama (1) adalah *Long Short-Term Memory* (LSTM), pemilihan ini didasarkan oleh beberapa pertimbangan. Algoritma ini telah

menjadi sorotan utama dalam komunitas *deep learning* dan mendapatkan beberapa pengakuan pada penelitian terdahulu atas perannya dalam model prediksi. Efisiensinya dalam pengolahan dataset *time series* dan kemampuannya dalam mencerna dan mengolah informasi data urutan jangka panjang juga patut diapresiasi. Kelebihan dari algoritma ini juga dapat dilihat dari Tabel 3.2, yang menjelaskan secara rinci performa dari LSTM pada beberapa penelitian terdahulu. Algoritma ini secara keseluruhan sangat bermanfaat untuk mengatasi dependensi jangka panjang yang terdapat dalam data dan membuatnya menjadi pilihan pertama yang akan diimplementasikan dalam pembangunan model prediksi harga *cryptocurrency*.

Selanjutnya, pilihan kedua (2) yang akan digunakan dalam penelitian ini, nantinya akan dievaluasi dan dibandingkan performanya dengan algoritma pilihan pertama (1) adalah *Convolutional Neural Network* (CNN). Algoritma ini tidak hanya terkenal dikarenakan oleh kemampuannya dalam menangkap pola pada gambar, namun juga terbukti dapat mengidentifikasi pola dan tren yang terdapat dalam dataset *time series*. CNN memiliki beberapa kelebihan tambahan, yaitu kemampuan algoritma untuk diadaptasikan dengan analisis *time series*, karena keefisienan model dalam menangkap pola pada data *time series*. Selanjutnya algoritma ini juga memiliki lapisan khusus dalam konfigurasi yang bekerja untuk mengurangi jumlah parameter, dengan demikian menghasilkan model yang ringkas, namun tetap efektif. Mengingat kembali bahwa penelitian bertujuan pada optimasi, CNN juga memberikan kelebihan dalam waktu pemrosesan dan skalabilitas (ranah komputasi), dengan demikian penggunaan algoritma sangat efektif dalam memprediksi dinamika pasar yang kerap berubah [68]. Adanya beberapa dukungan oleh hasil penelitian terdahulu dan pada bagian pendahuluan juga menyoroti potensi signifikan yang dapat diberikan oleh algoritma CNN dalam memprediksi harga *cryptocurrency*.

Meskipun terdapat beberapa penelitian yang mengkombinasikan LSTM dan CNN, karena memiliki potensi untuk dikombinasikan. Namun penting

untuk mengevaluasi dan membandingkan kedua model agar mendapatkan algoritma yang paling optimal dan efektif. Adapun penelitian terdahulu yang mendukung perbandingan antara kedua algoritma tersebut, yaitu penelitian yang dilakukan oleh Purnama et al. Dalam penelitian terdapat perbandingan antara algoritma LSTM dan CNN dalam prediksi data *time series* untuk sektor pariwisata dengan dataset pariwisata dan hasil yang menjadi lebih unggul pada dataset pariwisata adalah algoritma LSTM, karena kekuatannya dalam mencerna dan mengelola data urutan jangka panjang [22]. Sebaliknya, terdapat penelitian yang mengintegrasikan kedua algoritma *deep learning* menjadi satu, seperti penelitian Wenjie Lu et al., yang memprediksi harga saham dengan menggunakan kombinasi CNN-LSTM, dimana tingkat keoptimalan model dan efektivitas dilihat dari metrik MAE dan RMSE yang sangat memuaskan [23]. Penggunaan kedua algoritma ini dengan cara perbandingan maupun juga kombinasi pada penelitian terdahulu, menunjukkan bahwa masing-masing algoritma memiliki kelebihan tersendiri. Performa dari model juga dapat mencapai hasil yang signifikan, adanya peningkatan keoptimalan dan efektivitas model prediktif jika dikonfigurasi dengan tepat. Evaluasi antara kedua algoritma *deep learning* pilihan, yaitu (1) LSTM dan (2) CNN menjadi patokan dalam tahapan awal yang akan dilakukan agar dapat membuat model *hybrid* berbasis *machine learning* dan *deep learning* yang optimal dalam prediksi harga *cryptocurrency*.

3.2.3.2 Pilihan *Machine Learning* Model

Tabel 3. 3 Perbandingan Model XGBoost dan SVM

Indikator	XGBoost [43], [69], [70], [71]	Support Vector Machine (SVM) [72], [73]
Kelebihan	<ul style="list-style-type: none"> • Dapat digunakan untuk masalah regresi maupun klasifikasi • Umumnya memberikan hasil performa yang baik • Efisien, cepat, dan dirancang untuk menjadi <i>scalable</i> untuk data yang besar • Menyediakan berbagai macam pilihan parameter untuk penggunaan <i>fine-tuning</i> 	<ul style="list-style-type: none"> • Digunakan untuk kasus dimana data memiliki dimensi yang luas • Efisien dalam segi memori • Fleksibel dikarenakan memiliki kemampuan untuk menggunakan kernel (dapat menangani berbagai jenis data)
Kekurangan	<ul style="list-style-type: none"> • Menghabiskan waktu yang signifikan saat melakukan <i>training</i> • Efektivitas model tergantung pada <i>tuning</i> yang dilakukan 	<ul style="list-style-type: none"> • Sulit untuk melakukan modifikasi pada kernel • Tidak dianjurkan untuk menggunakan data yang besar • Susah untuk digunakan pada data yang memiliki noise besar

Tahapan selanjutnya dalam penelitian adalah integrasi algoritma *deep learning* terbaik dengan dua algoritma *machine learning* pilihan. Pemilihan algoritma *machine learning*, berdasarkan keunggulan yang ditawarkan serta reputasi kedua algoritma *machine learning* dalam meningkatkan keoptimalan dan keefisienan pembangunan model prediktif yang akan memproses data historis harga *cryptocurrency* yang berskala makro. Algoritma pilihan pertama (1) yang akan diintegrasikan bersama model *deep learning* terbaik, adalah model *eXtreme Gradient Boosting* (XGBoost). Keunggulan XGBoost menjadi alasan utama pemilihan algoritma ini dalam penelitian, keunggulan tersebut dapat dilihat dari Tabel 3.4 dan bagian pendahuluan. Efektivitas XGBoost dalam menghadapi *overfitting* sangat krusial dalam penelitian, dimana *overfitting* dalam penelitian ini sangat rentan karena fluktuasi harga *cryptocurrency* dan data *time series* yang kompleks. Adanya dukungan dari penelitian terdahulu

yang menggabungkan algoritma *deep learning* dengan XGBoost memberikan pemahaman terkait potensi integrasi yang dapat diadopsikan dalam penelitian. Kelebihan yang ditawarkan oleh XGBoost, tidak hanya meningkatkan efisiensi dalam pembangunan model, namun juga membawakan fleksibilitas dan optimalisasi proses penelitian secara menyeluruh. Salah satu penelitian yang sangat mendorong pemilihan algoritma ini adalah penelitian dari Nabipur et al., yang menjelaskan kemampuan XGBoost dalam mengatasi *overfitting*, mengelola data yang belum jelas (*null, missing values*), serta penelitian juga mengungkapkan algoritma XGBoost lebih unggul dibanding dengan metode berbasis pohon lainnya. Dengan demikian, XGBoost dijadikan pilihan pertama untuk diintegrasikan dengan model *deep learning* terbaik.

Algoritma *machine learning* pilihan kedua (2) dalam penelitian, adalah LightGBM. Keputusan untuk memilih LightGBM diambil berdasarkan keunggulan dari algoritma dan dukungan dari hasil penelitian terdahulu. LightGBM bekerja dengan kecepatan tinggi dan efisien terhadap penggunaan memori (pengurangan memori), tanpa mengorbankan keoptimalan dan akurasi dari model prediksi. Kemampuan LightGBM dalam mengurangi kompleksitas juga membuat algoritma menjadi salah satu pilihan. LightGBM memiliki potensi untuk berdampak jika dikombinasikan dengan algoritma *deep learning* terbaik. Penelitian dari Mulia et al., paralel dengan penelitian yang akan dilakukan, dimana terdapat proses evaluasi dan perbandingan dari algoritma LightGBM dengan XGBoost. Hasil yang didapatkan menunjukkan performa yang lebih akurat dan efisien dari algoritma LightGBM. Penelitian ini mendukung tahapan evaluasi dan perbandingan antara kedua algoritma dengan memberikan pemahaman mendalam dan informasi tambahan mengenai keunggulan dari kedua algoritma secara tunggal.

Dengan keunggulan yang dimiliki oleh kedua algoritma *machine learning*, penting untuk melakukan evaluasi dan perbandingan dengan *input* dataset harga historis *cryptocurrency* agar dapat mendapatkan algoritma yang optimal dan efektif. Tahapan terakhir penelitian adalah ini berfokus

pada evaluasi dan perbandingan dari integrasi algoritma *deep learning*, baik LSTM atau CNN dengan XGBoost dan juga dengan LightGBM. Hasil akhir berupa model *hybrid* terbaik dalam prediksi *cryptocurrency* dengan tingkat keoptimalan dan akurasi yang tinggi. Dengan demikian, diharapkan kombinasi dari *machine learning* dan *deep learning* dapat menghasilkan keoptimalan dan efektivitas model *hybrid* yang berguna dalam memprediksi harga *cryptocurrency* dan berkontribusi terhadap perkembangan teoritis dan praktik analisis model *hybrid* berbasis *machine learning* dan *deep learning*.

3.3 Teknik Pengumpulan Data

Terdapat *library yfinance* yang juga digunakan dalam memperoleh data dalam *Jupyter Notebook*. Dengan menggunakan *library yfinance*, data yang terdapat pada Yahoo Finance (*Cryptocurrency*) dapat *extract* langsung ke dalam platform *open-source* yang digunakan dan dapat disesuaikan jangka waktu data yang diperlukan.

3.3.1 Populasi dan Sampel

Penelitian mengimplementasikan teknik *purposive sampling* dalam proses mengumpulkan data. Menurut Sugiono, teknik pengambilan sampel *purposive sampling* mengutamakan sampel pilihan yang sesuai spesifikasi/pertimbangan tertentu [74]. Terdapat 3 jenis *cryptocurrency* yang digunakan dalam penelitian, yaitu *cryptocurrency* dengan peringkat kapitalisasi pasar yang mendominasi (peringkat 1 hingga 3) di tahun 2024. Data yang digunakan didapatkan pada *website* coinmarketcap.com dan dilihat dari ranking posisi kapitalisasi pasar. Berikut merupakan jenis *cryptocurrency* yang akan digunakan.

Tabel 3. 4 Jenis Cryptocurrency

Jangka Waktu Data	Rank	Nama Cryptocurrency	Singkatan
-------------------	------	---------------------	-----------

1 Januari 2018 – 31 Desember 2023	1	Bitcoin	BTC
	2	Ethereum	ETH
	3	Tether	USDT

3.3.2 Periode Pengambilan Data

Proses mengimpor data dari Yahoo Finance dilakukan pada bulan Maret 2024, sedangkan data historis yang diimpor merupakan data dari Januari 2018 hingga bulan Desember 2023. Data yang digunakan merupakan data asli dari Yahoo Finance yang bergerak secara fluktuatif (*daily*). Pengambilan data akan dilakukan kepada jenis *cryptocurrency* dengan posisi kapitalisasi pasar peringkat 1 hingga 3 pada *website* coinmarketcap.com.

Dasar pemilihan data *cryptocurrency* dari 1 Januari 2018 hingga 31 Desember 2023 karena fluktuasi harga yang lebih signifikan dibandingkan sebelum 2018. Gambar 3.4 menunjukkan harga historis Bitcoin, Ethereum, dan Tether sebelum 2018, harga relatif stabil dan kurang berfluktuasi, sementara setelah 2018, terjadi peningkatan dan penurunan harga yang signifikan. Periode setelah 2018 lebih representatif untuk analisis model prediktif karena data yang lebih dinamis dan variabel, yang diperlukan untuk melatih algoritma secara efektif. Menggunakan data setelah 2018 meningkatkan akurasi dan efektivitas prediksi harga *cryptocurrency*.

U N I V E R S I T A S
M U L T I M E D I A
N U S A N T A R A



Gambar 3. 4 Bukti Fluktuasi Cryptocurrency

Sumber : [10]

3.4 Variabel Penelitian

1. Variabel Independen

Variabel independen merupakan variabel yang menjadi acuan, dapat diukur dan menjadi sebab mengapa terdapatnya perubahan pada variabel dependen (terikat) [75]. Variabel independen yang terdapat dalam penelitian dibagi menjadi dua bagian, yaitu sebagai berikut:

- A. *Cryptocurrency Historical Data* : Variabel yang digunakan sebagai dasar untuk membuat prediksi pergerakan harga *cryptocurrency*.
- B. Algoritma: Penelitian ini menggunakan 4 algoritma, yaitu 2 algoritma *deep learning*, *Convolutional Neural Networks* (CNN) dan *Long Short-Term Memory Network* (LSTM), serta 2 algoritma *machine learning*, yaitu *Extreme Gradient Boosting* (XGBoost) dan *Light Gradient Boosting Machine* (LightGBM).

2. Variabel Dependen

Variabel dependen merupakan variabel yang akan diprediksikan dalam sebuah penelitian ataupun juga variabel yang dapat dipengaruhi oleh variabel-variabel lainnya dalam penelitian. Saat mencari hubungan antara variabel dependen dengan satu atau lebih variabel independen, umumnya analisis regresi akan digunakan [76]. Dalam penelitian ini variabel dependen merupakan hasil prediksi yang diperoleh menggunakan *hybrid* model yang telah dikembangkan, dimana model ini menggunakan variabel-variabel independent agar dapat beroperasi.

3.5 Teknik Analisis Data

Terdapat bahasa pemrograman yang digunakan sebagai *tools* penelitian ini, agar dapat mendukung proses *data mining* dan perancangan yang dilakukan. Diantara berbagai macam bahasa pemrograman yang dapat digunakan dalam penelitian, telah diperkecil menjadi 2, yaitu Python dan R. Berikut merupakan perbandingan antara bahasa pemrograman yang digunakan dalam penelitian ini:

Tabel 3. 5 Perbandingan Bahasa Pemrograman

Indikator	Python [77], [78]	R [79]
Kelebihan	<ul style="list-style-type: none"> • Mudah untuk diaplikasikan dan dipelajari • Fleksibel terhadap perubahan dan dapat digunakan pada berbagai sistem operasi • Memiliki berbagai macam <i>library</i> • Dapat menggunakan berbagai macam IDE, seperti Jupyter Notebook, Spyder. 	<ul style="list-style-type: none"> • Memiliki sifat <i>open source</i> • Adanya berbagai macam <i>package & functions</i> • Dapat dengan mudah diintegrasikan dalam bahasa pemrograman lainnya
Kekurangan	<ul style="list-style-type: none"> • Kecepatan yang relative lambat dibandingkan dengan bahasa lainnya • Diperlukannya <i>memory space</i> yang besar • Adanya batasan pada <i>database access</i> 	<ul style="list-style-type: none"> • Sulit untuk dipelajari • Adanya <i>command</i> yang eksekif • Kompleks • Data yang disimpan, masuk di memori fisik • Eksekusi yang lama • IDE (Integrated Development Environment) pada R adalah Rstudio

Dapat dilihat dari Tabel 3.5, bahwa bahasa pemrograman Python dan R memiliki kekurangan dan kelebihan masing-masing. Namun yang lebih unggul dan akan digunakan dalam penelitian ini adalah bahasa pemrograman Python. Bahasa pemrograman ini dapat dikatakan sebagai bahasa pemrograman yang lebih universal dan dapat lebih mudah untuk dipahami. Python juga tidak hanya lebih simple, namun juga fleksibel dan memiliki sifat *multipurpose*. Terakhir, penting untuk mengingat kembali bahwa penelitian ini menggunakan *Jupyter Notebook*. Hal ini memungkinkan peneliti untuk secara maksimal memanfaatkan *library* yang tersedia pada *open-source* ini. Pilihan yang unggul dan diutamakan dalam penelitian ini adalah bahasa pemrograman Python.

Tabel 3. 6 Perbandingan Strategy Pembuatan Model *Hybrid*

Indikator	<i>Feature Extraction</i> [80], [81], [82]	<i>Boosting</i> [83], [84]
Kelebihan	<ul style="list-style-type: none"> • Meningkatkan akurasi analisis dan prediksi model, jika diterapkan untuk <i>hybrid</i> model • Memungkinkan adanya transformasi ulasan yang bermakna • Meningkatkan kemampuan generalisasi model terhadap jenis data yang berbeda • Mengatasi keterbatasan metode tunggal 	<ul style="list-style-type: none"> • Mencegah <i>overfitting</i> • Dapat meningkatkan hasil akurasi prediksi • Fleksibilitas terhadap penggunaan tugas regresi atau klasifikasi
Kekurangan	<ul style="list-style-type: none"> • Lebih kompleks • Sumber daya yang besar • Kurang generalisasi pada data baru 	<ul style="list-style-type: none"> • Sumber daya yang besar • <i>Overfitting</i> dapat terjadi dengan adanya <i>outlier</i> atau <i>noise</i> pada data • Lebih memakan waktu, karena adanya pembelajaran secara berurutan

Dari Tabel 3.6, terdapat sejumlah informasi yang dapat diperoleh mengenai strategi pembuatan model *hybrid*. Pertama, terdapat strategi Boosting yang dapat meningkatkan akurasi model dengan cara kerja *ensemble* (gabungan antara model yang memiliki kekurangan). Kedua, adalah strategi *Feature Extraction* yang mengambil representasi dari data model *deep learning* dan dijadikan sebagai *input* untuk model lainnya. Melihat dari spesifikasi dari kedua teknik dan penjelasan cara kerja model, peneliti memilih strategi *Feature Extraction* dalam perancangan model *hybrid*. Pemilihan ini juga didukung dengan hasil dan informasi yang diperoleh dari penelitian Elavarasan et al., yang membuktikan bahwa proses *Feature Extraction* menghasilkan performa yang lebih unggul dibandingkan dengan proses lainnya [85]. Penelitian Kaur et al., juga menjelaskan bahwa strategi *Feature Extraction* memungkinkan analisa yang efisien, andal dan kuat, jika

diimplementasikan dalam proses pembuatan model *hybrid*. Oleh karena itu, ini menjadi beberapa alasan atas pemilihan *Feature Extraction* sebagai strategi yang akan diadopsikan dalam penelitian.

3.6 Teknik Tuning Model

3.6.1 Tahap I: *Hyperband Tuning - Deep Learning*

Penelitian menggunakan teknik *Hyperband Tuning* pada tahapan awal penelitian, yaitu dalam pengembangan model prediksi harga *cryptocurrency* dengan algoritma CNN dan LSTM. Keunggulan dari teknik *Hyperband* adalah kemampuannya dalam mengalokasikan sumber daya (secara adaptif) dengan pendekatan berbasis *bandit* dan juga penggunaan *early stopping*. Dengan adanya penyaringan konfigurasi *hyperparameter* secara bertahap, *Hyperband* memungkinkan konfigurasi yang optimal untuk diteruskan sementara yang lain akan dihentikan terlebih dahulu, sehingga teknik dapat sangat menghemat sumber daya. Fokus utama dalam menentukan dan menemukan *hyperparameter* yang akan memberikan performa terbaik dengan meminimalisir waktu komputasi dan memaksimalkan performa model merupakan alasan mengapa teknik *Hyperband* digunakan pada tahapan awal pengembangan model *deep learning*.

3.6.2 Tahap II: *Optuna Tuning - Machine Learning*

Optuna menjadi teknik pilihan yang digunakan saat melakukan *tuning* pada tahap selanjut. *Tuning* dilakukan kepada kedua model *machine learning* (XGBoost dan LightGBM) yang menggunakan *input* dari model *deep learning* terbaik untuk dijadikan sebagai *hybrid* model. Teknik *Optuna* dipilih karena fleksibilitasnya dan kemampuan modular dalam optimasi *hyperparameter* seperti pada algoritma XGBoost dan LightGBM. Dengan adanya API *define-by-run*, teknik ini dapat mempermudah integrasi selama proses pengembangan. Pendekatan selektif yang adaptif dan keefisienan teknik dalam mengeliminasi *trial* yang kiranya tidak dapat dilanjutkan, kembali membawakan keoptimalan secara sumber daya dan waktu. Dengan gabungan optimasi *Gradien* dan *Bayesian Optimization*, *Optuna* secara cepat dapat menghasilkan konfigurasi model yang optimal dan efektif. Bahkan dapat secara keseluruhan berjalan lebih cepat

dibandingkan pada model *deep learning*. Kemampuan *Optuna* dalam menyesuaikan konfigurasi *hyperparameter* secara selektif dan dinamis, membuat model ini menjadi teknik yang ideal dalam penelitian.

Penggunaan *Hyperband* pada Tahap I memanfaatkan keefisienan dalam menangani model *deep learning*, yang membutuhkan sumber daya yang signifikan dengan durasi pelatihan yang panjang. Dengan mengimplementasikan pendekatan berbasis *bandit*, peneliti dapat mengoptimalkan proses *tuning* dengan secara efisien meminimalisir konfigurasi yang tidak layak ataupun tidak cocok. Kerangka Tahap I dapat dilihat pada Gambar 3.2. Sementara itu, *Optuna* lebih tepat untuk digunakan dalam Tahap II, karena sangat adaptif dan kemampuannya dalam mengeliminasi *trial* yang tidak produktif, menjadi alasan mengapa teknik itu tepat untuk menyempurnakan model *machine learning* (XGBoost dan LightGBM). Kerangka Tahap II terdapat pada Gambar 3.3, yang sebelumnya telah dijelaskan. Dengan kecepatan yang tinggi, teknik *Optuna* dapat mempermudah proses pengaturan konfigurasi. Kedua teknik, *Hyperband* dan *Optuna*, dipilih dalam penelitian karena kemampuannya yang adaptif dalam menyesuaikan dan mengoptimalkan model. Integrasi kedua teknik *tuning* diharapkan mampu membangun model dengan kecepatan dan efisiensi yang tinggi, serta adaptif dalam optimasi model. Oleh karena itu, model *hybrid* yang mengkombinasikan *machine learning* dan *deep learning* diharapkan dapat mencapai performa terbaik dengan implementasi teknik *tuning* pilihan.