

BAB 2 LANDASAN TEORI

2.1 Cryptocurrency

Cryptocurrency adalah aset digital yang berfungsi sebagai alat tukar, dengan catatan kepemilikan untuk setiap koin yang tercatat dalam buku besar dan laporan keuangan yang disimpan dalam basis data yang terdiri dari serangkaian metode kriptografi. Terdapat jenis aset digital yang tidak berwujud atau mata uang digital dari *Cryptocurrency* yang menggunakan enkripsi sangat canggih disebut kriptografi untuk mengamankan dan memverifikasi transaksi. *Cryptocurrency* tidak dikendalikan oleh otoritas pusat seperti bank, negara atau perusahaan tertentu [3]. Tetapi menggunakan server yang terdesentralisasi yang berarti data, transaksi, atau informasi tidak diatur oleh satu server atau institusi, melainkan oleh jaringan komputer yang saling terhubung secara *peer-to-peer*. [4]

Cryptocurrency memungkinkan transaksi keuangan yang lebih cepat, lebih aman, dan lebih transparan tanpa memerlukan perantara seperti bank. Dengan menggunakan teknologi *Blockchain* yang menggunakan prinsip kriptografi yang kuat untuk melindungi integritas data dan mencegah perubahan dan pemalsuan data. Berikut dua mata uang crypto yang terkenal didunia, diantaranya:

- BTC
Mata uang digital BTC merupakan mata uang digital *Cryptocurrency* pertama yang memiliki kapitalis pasar terbesar yang pertama kali diluncurkan oleh Satoshi Nakamoto pada tahun 2008 sebagai mata uang berbasis internet. BTC awal mula diluncurkan untuk menjadi alat pembayaran yang tidak bergantung pada pihak perantara seperti bank, tetapi menggunakan *Blockchain* [21].
- ETH
Ethereum adalah *Blockchain* yang dapat diprogram dengan koin aslinya sendiri, *Ether* [22]. Ethereum adalah mata uang kripto terdesentralisasi dengan simbol ETH. Ethereum tidak dikendalikan oleh pemerintah atau organisasi mana pun dan dapat digunakan untuk investasi maupun pembayaran. ETH adalah *Cryptocurrency* terbesar kedua, setelah BTC. *Blockchain* Ethereum adalah platform komputasi terdistribusi sumber terbuka yang menunjukkan nilai dari Kontrak Pintar (Smart Contract). Seseorang dapat dengan mudah membuat aplikasi terdesentralisasi dalam skala besar

dan mendapatkan manfaat dari distribusi yang disediakan oleh teknologi *Blockchain* [23].

2.2 *Time Series Data*

Serangkaian pengamatan data yang telah disusun dalam waktu dengan jarak yang merata disebut *time series*. Data dari *time series* dapat dikumpulkan setiap hari, mingguan, bahkan bulanan kemudian dari data tersebut dapat melihat tren yang diidentifikasi dari data yang telah dikumpulkan. Secara keseluruhan, terdapat tiga model dari *time series*, diantaranya model tren, model musiman, dan model siklus [24]. Teknik prediksi *time series* dibagi dalam dua kategori yaitu model prediksi berbasis matematika serta statistik dan model prediksi berbasis kecerdasan buatan. Pertama, model prediksi berbasis matematika dan statistik, contohnya exponential smoothing, moving average, ARIMA, dan regresi. Kedua, model prediksi berbasis kecerdasan buatan, contohnya simulated annealing, klasifikasi, neural network, algoritma genetika, model hybrid, dan genetic programming [25].

2.3 *Machine Learning*

Pendekatan kecerdasan buatan (AI) yang bertujuan untuk menggantikan serta meniru peran manusia dalam melakukan suatu aktivitas maupun untuk memecahkan suatu masalah merupakan *machine learning*. Kata lain *machine learning* merupakan mesin yang dirancang untuk belajar yang berfungsi tanpa instruksi apa pun dari penggunanya. *Machine learning* secara umum diklasifikasikan menjadi tiga jenis, diantaranya *reinforcement learning*, *unsupervised learning*, dan *supervised learning* [26].

2.4 *Hyperparameter Tuning*

Hyperparameter tuning adalah proses menemukan kumpulan *hyperparameter* optimal yang menghasilkan performa terbaik untuk model Machine Learning [27]. *Hyperparameter* adalah parameter yang nilainya ditetapkan sebelum proses pembelajaran dimulai. *Hyperparameter* digunakan untuk mengelola berbagai aspek pembelajaran mesin yang memiliki dampak signifikan pada kinerja dan model yang dihasilkan [28].

2.5 Augmented Dickey-Fuller (ADF)

Uji Augmented Dickey-Fuller (ADF) mempertimbangkan adanya autokorelasi dalam residual dengan menambahkan variabel independen berupa lag dari perbedaan data. Uji ini menggunakan pendekatan statistik nonparametrik untuk mendeteksi autokorelasi dalam residual tanpa melibatkan lag perbedaan sebagai variabel penjelas. Dalam pelaksanaannya, uji ADF menghitung nilai statistik t dari koefisien γ dengan pendekatan khusus yang memperhitungkan derajat kebebasan berdasarkan jumlah observasi dan tingkat signifikansi tertentu. Namun, nilai kritis yang digunakan tidak berasal dari tabel distribusi t standar, karena hal ini dapat menyebabkan penolakan hipotesis nol secara berlebihan (over-rejection), sehingga data yang sebenarnya tidak stasioner dapat disimpulkan sebagai stasioner. Proses untuk menentukan apakah data bersifat stasioner dilakukan dengan membandingkan nilai statistik ADF dengan nilai kritisnya. Jika nilai absolut statistik ADF lebih besar daripada nilai kritis, maka data dianggap stasioner. Sebaliknya, jika nilai absolut statistik ADF lebih kecil dari nilai kritis, maka data dinyatakan tidak stasioner [29].

```
1 statsmodels.tsa.stattools.adfuller(  
2     x,  
3     maxlag=None,  
4     regression='c',  
5     autolag='AIC',  
6     store=False,  
7     regresults=False  
8 )
```

Parameter x bertipe array-like merupakan kumpulan data yang akan diuji. Parameter $maxlag$ None, int mengacu pada lag maksimum yang dimasukkan dalam pengujian. Jika diatur ke None, nilai defaultnya adalah $12 \cdot (\text{nobs}/100)^{1/4}$. Parameter $regression$ berisikan "c", "ct", "ctt", "n" menentukan jenis konstanta dan tren yang akan dimasukkan dalam regresi, dengan opsi sebagai berikut: "c" hanya memasukkan konstanta (default), "ct" memasukkan konstanta dan tren, "ctt" mencakup konstanta serta tren linier dan kuadrat, sementara "n" berarti tanpa konstanta maupun tren. Parameter $autolag$ berisikan "AIC", "BIC", "t-stat", None adalah metode untuk secara otomatis menentukan panjang lag dari rentang 0 hingga $maxlag$. Jika memilih "AIC" (default) atau "BIC", panjang lag dipilih berdasarkan kriteria informasi yang diminimalkan. Jika menggunakan "t-stat", pemilihan dimulai dari $maxlag$ dan mengurangi lag hingga t-statistik pada

lag terakhir signifikan pada tingkat uji 5%. Jika diatur ke None, panjang lag akan tetap sesuai dengan nilai maxlag. Parameter store adalah nilai boolean (True atau False). Jika True, hasil pengujian akan mencakup objek tambahan selain statistik ADF. Nilai default parameter ini adalah False. Sedangkan, parameter regresults juga berupa boolean yang menentukan apakah hasil lengkap dari regresi akan ditampilkan (True) atau tidak (False). Nilai defaultnya adalah False [30].

2.6 P-value

P-Value merupakan probabilitas yang menunjukkan seberapa besar kemungkinan mendapatkan hasil yang sama atau lebih ekstrem dari hasil yang diamati, dengan asumsi bahwa tidak ada pengaruh atau perbedaan (hipotesis nol). Nilai ini mengindikasikan peluang bahwa perbedaan yang teramati antara kelompok terjadi secara kebetulan semata [31]. CDF, atau Cumulative Distribution Function, adalah fungsi dalam statistik yang menunjukkan probabilitas bahwa suatu variabel acak X memiliki nilai kurang dari atau sama dengan x . Persamaan 2.1 merupakan rumus matematis untuk mencari P-value [32].

$$P = 2 \cdot (1 - \text{CDF}(|t_s|)) \quad (2.1)$$

2.7 ARIMA

ARIMA (*Auto Regressive Integrated Moving Average*) merupakan model yang menggunakan nilai variable dependen masa lalu dan saat ini untuk membuat prediksi jangka pendek yang lebih akurat [33]. Metode ARIMA memiliki tiga proses utama, diantaranya pemeriksaan diagnostik, identifikasi, dan prediksi [34] [35].

1. Proses pemeriksaan diagnostik melibatkan pengendalian stasioneritas data deret waktu. Hal ini penting untuk membangun model ARIMA yang berfungsi untuk membuat deret waktu non-stasioner menjadi stasioner, diferensiasi (d) yang sesuai dilakukan dan stabilitasnya diuji kembali. Proses ini berlanjut hingga diperoleh deret stasioner. Setelah beberapa operasi diferensiasi, parameter integral model ARIMA diatur ke (d).
2. Proses identifikasi dilakukan dengan menggunakan data stasioner untuk menentukan parameter operasi *Autoregressive* (AR) dan *Moving Average* (MA).

3. Prediksi dibuat menggunakan model ARIMA. Ini didefinisikan sebagai ARIMA (p, d, q). Dimana p adalah derajat model AR, d adalah derajat differencing, dan q adalah derajat MA [34]. Persamaan 2.1 merupakan rumus matematis model ARIMA.

$$y_t = \alpha_1 w_{t-1} + \alpha_2 w_{t-2} + \dots + \alpha_p w_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}. \quad (2.2)$$

- (a) y_t : data riil yang dilinierisasi pada waktu ke- t .
- (b) t : waktu.
- (c) $\alpha_1, \alpha_2, \dots, \alpha_p$: parameter *Auto Regressive* (AR).
- (d) $w_{t-1}, w_{t-2}, \dots, w_{t-p}$: data yang diobservasi.
- (e) ε_t : residual acak yang tidak diketahui (error).
- (f) $\theta_1, \theta_2, \dots, \theta_q$: rata-rata bergerak (*moving average*).
- (g) $\varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_{t-q}$: data error.

2.8 GARCH

Terdapat masalah heteroskedastisitas dalam variabilitas residual pada data deret waktu. Perubahan variabilitas residual ini disebabkan oleh pengaruh nilai residual dari periode sebelumnya. Berdasarkan analisis, diperkenalkan konsep heteroskedastisitas untuk membentuk model yang disebut *Autoregressive Conditional Heteroscedasticity* (ARCH) dengan ordo q. Istilah *conditional heteroscedasticity* merujuk pada fakta bahwa variabilitas residual pada suatu periode dipengaruhi oleh nilai residual dari periode sebelumnya. Variabilitas residual tidak hanya dipengaruhi oleh nilai residual dari periode sebelumnya, tetapi juga oleh variabilitas residual dari periode terdahulu. Pada tahun 1986, model ARCH diperluas dengan menambahkan unsur nilai residual dan variabilitas residual dari periode sebelumnya. Model yang dikembangkan ini dikenal sebagai *Generalized Autoregressive Conditional Heteroskedasticity* (GARCH) dengan ordo p dan q. Model ini disebut sebagai model GARCH(p,q) [17].

$$\sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2 \quad (2.3)$$

Model GARCH(p,q):

$$\sigma_t^2 = \omega + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^p \beta_i \sigma_{t-i}^2 \quad (2.4)$$

dimana:

- σ_t^2 : ragam residual
- ω : konstanta dengan syarat $\omega > 0$
- α_i : koefisien dari residual lampau dengan syarat $\alpha_i \geq 0, \quad i = 1, \dots, q$
- β_i : koefisien dari ragam residual lampau dengan syarat $\beta_i \geq 0, \quad i = 1, \dots, p$
- $\sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2$: sigma kuadrat residual periode yang lalu (lag q)
- $\sum_{i=1}^p \beta_i \sigma_{t-i}^2$: sigma ragam residual periode yang lalu (lag pt)

2.9 Hybrid ARIMA-GARCH

Model Hybrid ARIMA-GARCH dibangun melalui dua tahap. Tahap pertama, digunakan model ARIMA terbaik yang sesuai dengan data *time series* stasioner dan linier, dengan bagian data non-linier yang terkandung dalam residual model linier. Tahap kedua, digunakan model GARCH untuk menangkap pola residual non-linier. Model hybrid ini gabungan dari model ARIMA dan GARCH dengan pola residual non-linier [19]. Persamaan 2.4 merupakan rumus Hybrid ARIMA-GARCH [36].

$$\hat{Y}_t^{hybrid} = \hat{L}_t^{ARIMA} + E_t^{EGARCH} \quad (2.5)$$

2.10 Root Mean Square Error

Root Mean Square Error (RMSE) adalah metrik matematika yang diperoleh dari *Mean Square Error* (MSE) yang ditemukan saat mengevaluasi metode tertentu. RMSE adalah metrik yang umum digunakan dalam pemodelan prediktif. Metrik ini digunakan untuk mengevaluasi keakuratan prediksi model dengan membandingkan prediksi model dengan observasi terkait. RMSE menyediakan angka tunggal yang mengukur perbedaan keseluruhan antara prediksi model dan hasil aktual dengan menghitung akar kuadrat dari rata-rata selisih kuadrat antara nilai prediksi dan observasi. Untuk persamaan RMSE matematis dapat digunakan Persamaan (2.5) [37].

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Actual_i - Predicted_i)^2} \quad (2.6)$$

2.11 Mean Absolute Percentage Error

Mean Absolute Percentage Error (MAPE) adalah perbedaan absolut rata-rata antara nilai prediksi dan nilai aktual, yang dinyatakan sebagai persentase dari nilai sebenarnya. Penggunaan MAPE untuk mengevaluasi hasil prediksi memungkinkan Anda melihat seberapa akurat angka prediksi dan angka sebenarnya [38]. Nilai MAPE dapat dihitung menggunakan persamaan (2.4) [39].

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \frac{|Actual_i - Predicted_i|}{Actual_i} \quad (2.7)$$

Evaluasi MAPE didasarkan pada besar kecilnya nilai MAPE. Nilai MAPE yang kecil menunjukkan bahwa model prediksi tersebut sangat akurat, sedangkan nilai MAPE yang besar menunjukkan bahwa model prediksi tersebut kurang akurat. Kisaran nilai MAPE dapat dilihat pada Tabel 2.1 [40].

Tabel 2.1. Akurasi nilai MAPE

MAPE	Akurasi
<10%	akurasi prediksi sangat baik
10-20%	akurasi prediksi baik
20-50%	akurasi prediksi memadai
>50%	akurasi prediksi buruk

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