

BAB II TINJAUAN PUSTAKA

2.1 Penelitian Terkait

Beberapa Penelitian yang terkait dengan Cryptocurrency dan berbagai metode algoritma yang di pakai:

Tahun	Peneliti	Topik	Metode	Hasil	Simpulan
2020	Moch Farryz Rizkilloh, Sri Widiyanesti Manajemen Bisnis Telekomunikasi dan Informatika, Fakultas Ekonomi dan Bisnis, Universitas Telkom	Prediksi Harga Cryptocurrency Menggunakan Algoritma Long Short Term Memory (LSTM)	Long Short Term Memory (LSTM)	nilai RMSE yang didapatkan pada koin DOGE sebesar 0.0544 sedangkan nilai RMSE koin ADA sebesar 0.1607.	Penggunaan LSTM terbukti dapat digunakan untuk membangun model prediksi terhadap data time series karena LSTM memiliki kemampuan untuk mengingat dan menyimpan History data baik untuk jangka pendek maupun jangka Panjang. Hal ini dibuktikan dengan keberhasilan LSTM dalam memprediksi harga koin untuk hari selanjutnya. Selain itu hasil pengujian terhadap jenis koin DOGE dan ADA yang mendapatkan

					nilai RMSE yang cukup baik dimana nilai RMSE yang didapatkan pada koin DOGE sebesar 0.0544 sedangkan nilai RMSE koin ADA sebesar 0.1607. [5]
2022	Indriyanti , Nurul Ichsan, Haerul Fatah, Tri Wahyuni4, Erni Ermawati. Universitas Bina Sarana Informatika	IMPLEMENTASI ORANGE DATA MINING UNTUK PREDIKSI HARGA BITCOIN	K-Nearest Neighbor (K-NN) dan metode Support Vector Machine	Nilai RMSE dan MAE dari metode <i>K-Nearest Neighbor (K-NN)</i> sebesar 0.036 dan 0.027. Sedangkan metode <i>Support Vector Machine</i> memiliki nilai RMSE dan MAE sebesar 0.010 dan 0.008.	Hasil implementasi orange data mining untuk prediksi harga bitcoin dapat dilihat pada <i>widjet Test and Score</i> , dimana diperoleh hasil nilai RMSE dan MAE dari masing-masing model. Nilai RMSE dan MAE dari metode <i>K-Nearest Neighbor (K-NN)</i> sebesar 0.036 dan 0.027. Sedangkan metode <i>Support Vector Machine</i> memiliki nilai RMSE dan MAE sebesar 0.010 dan 0.008. [6]
2018	Bayu	Klasifikasi	Neighbor	hasil	Berdasarkan

	Laksana Yudha1 , Lailil Muflikhah2 , Randy Cahya Wihandika3. Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Brawijaya	Risiko Hipertensi Menggunakan Metode Neighbor Weighted K Nearest Neighbor (NWKNN)	Weighted K Nearest Neighbor (NWKNN)	akurasi terbaik pada K=10, E=4 dan data latih sebanyak 100 data dengan akurasi 88%.	hasil pengujian, metode NWKNN dapat melakukan klasifikasi risiko hipertensi dengan hasil akurasi terbaik pada K=10, E=4 dan data latih sebanyak 100 data dengan akurasi 88%. [7]
2020	1 Nur Fitriani Bintang Pradana, 2 Sri Lestanti 3Fakultas Tekonolgi Informasi Universitas Islam Balitar	APLIKASI PREDIKSI JANGKA PENDEK HARGA BITCOIN MENGGUNAKAN METODE ARIMA	Autoregressive Integrated Moving Average (ARIMA)	Rata-rata nilai MAPE yang dihasilkan adalah sebesar 0,84 dengan rentan nilai sebesar 1,34 untuk prediksi hari pertama (1 Juli 2020) dan 0,98 untuk prediksi hari ketujuh (7 Juli 2020)	Kesimpulan yang diperoleh dari penelitian ini yaitu ARIMA telah berhasil diterapkan pada aplikasi berbasis web untuk melakukan prediksi harga bitcoin untuk satu hingga tujuh hari kedepan dengan hasil yang baik. Beberapa model telah diuji dan model ARIMA (3,1,3) dapat melakukan prediksi harga bitcoin dengan tingkat akurasi yang baik. Rata-rata nilai MAPE yang dihasilkan adalah sebesar

					0,84 dengan rentan nilai sebesar 1,34 untuk prediksi hari pertama (1 Juli 2020) dan 0,98 untuk prediksi hari ketujuh (7 Juli 2020). Performa ARIMA semakin baik jika digunakan untuk prediksi jangka pendek, terutama untuk prediksi periode dua hari kedepan. [8]
2021	Siti Saadah 1 , Haifa Salsabila 2 1Informatics Faculty, Telkom University	Prediksi Harga Bitcoin Menggunakan Metode Random Forest	Random Forest	Performansi yang baik dengan nilai MAPE sebesar 1.50% atau dengan akurasi yang diperoleh sekitar 98% dengan menggunakan data acak.	Berdasarkan hasil penelitian, algoritma random forest dengan menggunakan atribut low, high dan price berhasil memprediksi harga bitcoin dengan performansi yang baik dengan nilai MAPE sebesar 1.50% atau dengan akurasi yang diperoleh sekitar 98% dengan menggunakan data acak. Perolehan akurasi untuk prediksi dengan data

					acak ini menunjukkan nilai yang setara yakni sekitar 95% - 98%. Sehingga mengacu kepada hasil prediksi ini dapat disimpulkan bahwa untuk memprediksi data yang memiliki karakteristik seperti bitcoin, yakni memiliki fluktuasi tinggi, random forest dikatakan berhasil memberikan fitting yang sesuai dengan data sesungguhnya. Akan tetapi, hasil ini berbanding terbalik pada saat memprediksi menggunakan data tidak acak. [9]
2018	Haerul Fatah1; Agus Subekti2 1Program Studi Ilmu Komputer STMIK Nusa Mandiri Jakarta	PREDIKSI HARGA CRYPTOCURRENCY DENGAN METODE K-NEAREST NEIGHBOURS	<i>KNearest Neighbors</i>	<i>KNN</i> dengan nilai K=3	Kesimpulan dari penelitian yang dilakukan yaitu: Dari hasil eksperimen metode <i>KNN (KNearest Neighbors)</i> pada 3 buah dataset dengan

					<p>menggunakan parameter nilai K dan <i>Nearest Neighbour Search Algorithm</i>, dapat disimpulkan model <i>KNN</i> yang memiliki akurasi paling baik adalah <i>KNN</i> dengan nilai $K=3$ dan <i>Nearest Neighbour Search Algorithm: Linear NN Search</i>. Semakin besar nilai K maka semakin besar pula nilai <i>Mean Absolute Error</i> dan nilai <i>Root Mean Squared Error</i> nya. Serta semakin banyak dataset/data <i>history</i> harga <i>Cryptocurrency</i> yang digunakan, maka semakin kecil nilai <i>Mean Absolute Error</i> dan nilai <i>Root Mean Squared Error</i> nya. [10]</p>
2020	Kinanti Dhea Larasati PROGRAM STUDI STATISTIKA	PREDIKSI HARGA BITCOIN BERDASARKAN INFORMASI	<i>Multivariate LSTM</i>	berdasarkan nilai MSE didapatkan bahwa model	Prediksi menggunakan <i>Multivariate LSTM</i> langkah pertama yang harus

	<p>FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM UNIVERSITAS ISLAM INDONESIA YOGYAKARTA</p>	<p>BLOCKCHAIN MENGGUNAKAN METODE <i>LONG-SHORT TERM MEMORY</i></p>		<p>dengan jumlah neuron 20 dan <i>epoch</i> 500 memiliki nilai MSE terkecil.</p> <p>tingkat akurasi sebesar 91,07% berdasarkan nilai MAPE.</p>	<p>dilakukan adalah normalisasi menggunakan <i>min-max scaling</i>. Langkah kedua yaitu membagi data menjadi data <i>training</i> dan data <i>testing</i>, dalam penelitian ini menggunakan perbandingan 80% data <i>training</i> dan 20% data <i>testing</i>. Langkah ketiga yaitu menentukan jumlah <i>neuron</i> dan <i>epoch</i>, pada penelitian ini menggunakan jumlah <i>neuron</i> 10, 20, 30, 40, dan 50, sedangkan jumlah <i>epoch</i> yang digunakan adalah 100, 500, dan 1000, berdasarkan nilai MSE didapatkan bahwa model dengan jumlah neuron 20 dan <i>epoch</i> 500 memiliki nilai MSE terkecil.</p> <p>Kemudian dilakukan prediksi dan</p>
--	---	--	--	--	--

					diperoleh bahwa nilai prediksi tidak jauh berbeda dari data aktual, dengan tingkat akurasi sebesar 91,07% berdasarkan nilai MAPE.
2020	Ngakan Nyoman Pandika Pinata, I Made Sukarsa, Ni Kadek Dwi Rusjyanthi. Program Studi Teknologi Informasi, Fakultas Teknik, Universitas Udayana	Prediksi Kecelakaan Lalu Lintas di Bali dengan XGBoost pada Python	Algoritma <i>XGBoost</i>	RMSE 21,69 untuk model peramalan jumlah kejadian, RMSE 4,92 untuk model peramalan jumlah orang meninggal dunia, RMSE 4,11 untuk model peramalan jumlah orang yang mengalami luka berat dan RMSE 77,24 untuk model peramalan jumlah orang yang mengalami luka ringan.	Berdasarkan hasil penelitian mengenai penerapan XGBoost (Xtreme Gradient Boosting) untuk meramalkan jumlah kejadian, jumlah orang meninggal dunia, jumlah orang yang mengalami luka ringan, dan luka berat pada setiap tahunnya menghasilkan nilai <i>error</i> yang cukup rendah. Nilai <i>error</i> dari 4 kategori masing-masing yaitu RMSE 21,69 untuk model peramalan jumlah kejadian, RMSE 4,92 untuk model peramalan

					jumlah orang meninggal dunia, RMSE 4,11 untuk model peramalan jumlah orang yang mengalami luka berat dan RMSE 77,24 untuk model peramalan jumlah orang yang mengalami luka ringan. [11]
2022	Sri Elina Herni Yulianti, Oni Soesanto, Yua na Sukmawaty. Program Studi Statistika, Universitas Lambung Mangkurat, Indonesia	Penerapan Metode Metode <i>Extreme Gradient Boosting</i> (XGBOOST) pada Klasifikasi Nasabah Kartu Kredit	Algoritma <i>XGBoost</i>	diperoleh akurasi model sebesar 83,42%, presisi sebesar 85,36%, <i>recall</i> sebesar 95,28%	Berdasarkan hasil penelitian pada pembahasan didapatkan kesimpulan bahwa Hasil klasifikasi menggunakan metode <i>XGBoost</i> dengan parameter yang <i>default</i> pada dataset nasabah pengguna kartu kredit menghasilkan model yang dikatakan cukup baik yaitu akurasi model sebesar 80,02% , untuk presisi sebesar 85,32%, <i>recall</i> sebesar 94,86% dan

					<p>dapat dikategorikan sebagai <i>good classification</i>. Untuk percobaan kedua menggunakan teknik optimasi yaitu proses <i>hyperparameter tuning</i> menggunakan 7 <i>hyperparameter</i> dengan memvalidasi data, maka didapatkan hasil <i>hyperparameter tuning</i> yang diperoleh akurasi model sebesar 83,42%, presisi sebesar 85,36%, <i>recall</i> sebesar 95,28% dan hasil klasifikasi termasuk kategori <i>good</i>. [12]</p>
2022	Beno Jange, Program Studi Komputerisasi Akuntansi, STMIK Dharmapala Riau, Pekanbaru	Prediksi Harga Saham Bank BCA Menggunakan XGBoost	Algoritma <i>XGBoost</i>	Melalui beberapa penyetelan parameter hiper pada metode XGBoost akan didapatkan prediksi yang lebih baik yaitu MAPE	Berdasarkan hasil pembahasan yang telah dilakukan maka dapat disimpulkan bahwa metode XGBoost cukup baik dalam memprediksi harga saham

				<p>sebesar 4.01 persen.</p>	<p>Bank BCA. Melalui beberapa penyetelan parameter hiper pada metode XGBoost akan didapatkan prediksi yang lebih baik yaitu MAPE sebesar 4.01 persen. Dapat dilihat bahwa pemakaian indikator teknikal dalam rekayasa fitur dan penyetelan parameter hiper cukup signifikan di dalam menurunkan nilai MAPE. Namun perlu diingat bahwa semakin banyak indikator teknikal yang digunakan dalam rekayasa fitur dan jumlah parameter yang digunakan dalam seleksi parameter hiper maka waktu untuk memprosesnya menjadi semakin lama. Jika dibandingkan dengan</p>
--	--	--	--	-----------------------------	--

				<p>penelitian sebelumnya menggunakan Prophet bisa dilihat bahwa prediksi XGBoost berhasil melampaui prediksi Prophet namun ada banyak faktor penyebabnya yaitu jumlah parameter yang digunakan dan bahwa Prophet tidak menggunakan indikator teknikal dalam melakukan rekayasa fitur. Untuk penelitian selanjutnya disarankan untuk menggunakan data historis harga saham yang lebih panjang, menggunakan lebih banyak parameter hiper; membandingkan metode XGBoost dengan metode peramalan saham lainnya (seperti ARIMA, Regresi Linear, LSTM, KNN);</p>
--	--	--	--	--

					menggunakan lebih banyak saham untuk dibandingkan; memecah data yang sebelumnya data latih (train) dan uji (test) menjadi data latih (train), validasi (validation) dan uji (test). [13]
--	--	--	--	--	---

2.2 Landasan Teori

2.2.1 Cryptocurrency

Cryptocurrency adalah bentuk mata uang digital yang dibuat melalui rangkaian kode yang disebut blockchain. Uang kripto dapat digunakan sebagai alat pembayaran yang transaksinya dilakukan secara virtual melalui internet. Kelebihan mata uang kripto dibandingkan dengan mata uang konvensional meliputi fleksibilitasnya di mana saja, transparansi, kecepatan, dan biaya transaksi yang rendah. [2]

Cryptocurrency juga mencakup serangkaian teknologi berbasis kriptografi dan algoritma yang menggunakan kode dan sandi matematis untuk menciptakan mata uang digital. Kriptografi memberikan tingkat keamanan yang tinggi dengan mengurangi risiko pemalsuan mata uang.

2.2.1.1 Fungsi Cryptocurrency

Berikut adalah beberapa peran cryptocurrency yang perlu diperhatikan:

1. Untuk Membeli Barang atau Jasa

Peran utama mata uang digital adalah sebagai alat pembayaran untuk membeli barang atau jasa. Saat ini, banyak toko atau pengecer yang menerima mata uang digital sebagai metode pembayaran untuk barang atau jasa yang mereka tawarkan. Contohnya, beberapa toko online seperti Overstock dan Newegg telah menerima pembayaran menggunakan mata uang digital. Selain itu, beberapa restoran, hotel, aplikasi perjalanan, dan bahkan beberapa universitas telah mulai menerima mata uang digital sebagai alat pembayaran. Namun, penting untuk dicatat bahwa tidak semua jenis mata uang digital diterima secara luas saat ini, hanya Bitcoin yang paling umum diterima untuk keperluan tersebut.

2. Sebagai Bentuk Investasi

Mata uang digital juga berfungsi sebagai bentuk investasi. Prinsip investasi dalam mata uang digital mirip dengan investasi pada umumnya, di mana Anda membelinya ketika harganya rendah dan menjualnya ketika harganya meningkat. Namun, investasi dalam mata uang digital ini juga memiliki risiko yang tinggi. Pada awalnya, banyak orang mendapat keuntungan besar dari investasi ini karena harga mata uang digital naik secara signifikan. Namun, pertumbuhan investasi ini telah melambat sejak beberapa tahun lalu.

3. Mining

Fungsi lain dari mata uang digital adalah mining. Mining merujuk pada proses memecahkan algoritma kriptografi untuk mendapatkan mata uang digital tersebut. Proses ini memerlukan daya komputasi yang signifikan, dan semakin besar daya komputasi yang digunakan, semakin besar peluang untuk memperoleh mata uang digital.

2.2.1.2 Cara Kerja Cryptocurrency

Cryptocurrency dapat dianggap sebagai alat pembayaran yang hampir setara dengan alat pembayaran uang fisik saat ini. Perbedaannya terletak pada ketiadaan pihak sentral yang mengatur jumlah peredaran mata uang tersebut. Ini karena cryptocurrency beroperasi tanpa otoritas pusat yang mengendalikan pasokan mata uang. Proses kerja cryptocurrency melibatkan pembuatan bukti kriptografi yang diverifikasi dan dicatat dalam teknologi blockchain setiap kali terjadi transaksi menggunakan cryptocurrency.

2.2.2 Data Mining

Data mining merupakan suatu teknik pengolahan data yang menemukan hubungan dalam data yang tidak diketahui penggunaannya dan menyajikannya dalam format yang

mudah dipahami sehingga dapat digunakan sebagai dasar pengambilan keputusan (Ridwan et al., 2013). Data mining dapat dikelompokkan berdasarkan tugas yang dilakukan seperti deskripsi, estimasi, prediksi, klasifikasi, klasterisasi, dan asosiasi (Muslehatin dkk., 2017)

Definisi umum dari data mining itu sendiri adalah proses pencarian pola-pola yang tersembunyi (*hidden pattern*) berupa pengetahuan (*knowledge*) yang tidak diketahui sebelumnya dari suatu sekumpulan data yang mana data tersebut dapat berada di dalam database, data warehouse, atau media penyimpanan informasi yang lain.

Data mining dilakukan dengan menggunakan alat khusus yang melakukan operasi data mining tertentu berdasarkan model analitis. Data mining adalah proses analisis data yang berfokus pada pencarian informasi tersembunyi dalam sejumlah besar data yang disimpan dalam operasi bisnis. Kemajuan luar biasa yang terus berlanjut dalam data mining didorong oleh beberapa faktor, termasuk:

1. Pertumbuhan kumpulan data yang pesat.
2. Menyimpan data dalam data warehouse data memberi seluruh perusahaan akses ke database terpercaya.
3. Akses data melalui web dan navigasi internet semakin meningkat.
4. Tekanan persaingan usaha untuk meningkatkan kekuatan pasar dalam globalisasi ekonomi.
5. Pengembangan teknologi perangkat lunak untuk memungkinkan penggunaan teknologi data mining.
6. Perkembangan Utama dalam Pengembangan Kemampuan Daya Komputasi

dan Media Penyimpanan (Rahmawati & Merlina, 2018)

Secara umum, teknik data mining dapat dibagi menjadi dua kategori: deskriptif dan prediktif. Deskriptif berarti data mining mencari pola yang dapat dipahami manusia yang menggambarkan karakteristik data. Sedangkan prediksi berarti menggunakan data mining untuk membuat model pengetahuan yang digunakan untuk membuat prediksi (Suyanto, 2017) Teknik data mining adalah:

1. Klasifikasi

Klasifikasi merupakan suatu proses penemuan sekumpulan model yang mendeskripsikan kelas data sehingga model tersebut dapat digunakan untuk memprediksi nilai kelas yang tidak diketahui pada objek untuk mendapatkan modelnya, kita perlu menganalisis data pelatihan, sedangkan data uji digunakan untuk menentukan tingkat akurasi dan model yang dibuat. klasifikasi dapat digunakan untuk memprediksi nama atau nilai objek data.

2. Clustering

Mengelompokkan data yang belum diketahui label kelasnya ke dalam sejumlah kelompok tertentu menurut ukuran kemiripannya.

3. Asosiasi

Tujuan metode ini adalah untuk membuat seperangkat aturan yang menggambarkan kumpulan data yang berkaitan erat satu sama lain.

4. Regresi

Regresi mirip dengan klasifikasi perbedaannya terletak pada atribut yang menghasilkan nilai kontinu.

5. Prediksi

Fungsi prediksi (prediksi) yang memprediksi kejadian yang akan diolah berdasarkan data historis yang ada.

6. Sequence Analysis

Tujuan metode ini adalah untuk mengenali pola dari data diskrit, misalnya untuk menemukan kelompok gen dengan tingkat ekspresi serupa.

7. Deviation Analysis

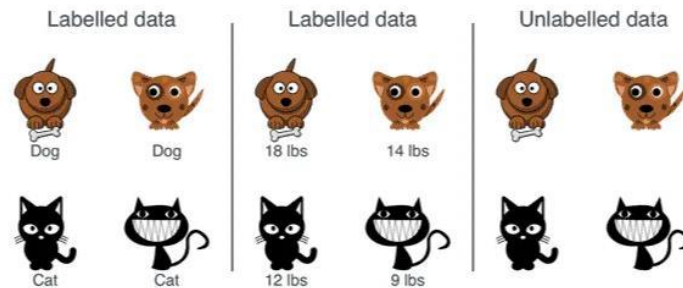
Tujuan metode ini adalah untuk mencari penyebab perbedaan antar data, ini biasa disebut deteksi outlier. Contohnya termasuk apakah pengguna kartu kredit ditipu dengan melihat data transaksi yang disimpan dalam database perusahaan.

2.2.3 Supervised dan Unsupervised Learning

Untuk lebih memahami Supervised dan unsupervised learning, supervised learning menggunakan data berlabel, sedangkan unsupervised learning menggunakan data tidak berlabel. Supervised learning digunakan untuk tugas klasifikasi dan regresi seperti deteksi objek, analisis prediktif, dan analisis sentiment sedangkan unsupervised learning digunakan untuk pengelompokan, asosiasi, dan pengurangan dimensi. Selanjutnya, juga perlu memahami jenis-jenis data tipe, data dibedakan menjadi dua jenis yaitu data berlabel (labeled data) dan data tidak berlabel (unlabeled data). data berlabel adalah data yang memiliki label berupa "tag" atau kelas, dan biasanya digunakan sebagai keluaran model. Sedangkan data tidak berlabel adalah data tidak berlabel yang digunakan sebagai keluaran pemodelan. misalnya dataset berupa kumpulan foto binatang. jika setiap

foto memiliki nama binatang, maka foto tersebut akan diberi label data penunjukan ini mengacu pada data kategorikal dan numerik.

Lihatlah diagram di bawah ini yang diambil dari Grokking Machine Learning.



Gambar 2.2.5.1 Ilustrasi Grokking Machine Learning

2.2.4 Forecasting Time Series

Forecasting adalah perkiraan nilai masa depan berdasarkan pola dalam suatu kumpulan data (Turban 2005). forecasting time series menggunakan data historis masa lalu untuk memprediksi apa yang akan terjadi di waktu yang akan datang. Time series adalah sekumpulan observasi yang diambil selama periode waktu tertentu, biasanya pada interval waktu yang besar (Turban 2005). forecasting dapat diterapkan ketika kondisi jika tersedianya informasi data historis, informasi tersebut dapat dikuantifikasi dalam bentuk numerik, dan beberapa aspek pola masa lalu diperkirakan akan terus berlanjut di masa yang akan datang [14].

Berikut adalah perhitungan forecasting time series:

Rumus umum metode Single Exponential Smoothing adalah sebagai berikut:

$$(S_{t+1} = \alpha X_t + (1 - \alpha) S_t)$$

S_{t+1} : periode peramalan (waktu hendak dilakukan peramalan)

X_t : data asli pada periode ke-t S_t : nilai peramalan pada periode ke-t

a : nilai perkiraan fluktuasi (diisi nilai antara 0 s /d 1)

Secara matematis, moving average dinyatakan dalam persamaan sebagai berikut:

$$F_{t+1} = \frac{X_t + X_{t-1} + \dots + X_{t-n+1}}{n}$$

F_{t+1} : nilai ramalan untuk periode waktu ke - t+1,

X_t : nilai aktual periode ke - t,

n : banyak data (ordo).

Sedangkan, untuk mengetahui tingkat kesalahan peramalan penjualan atau *Mean Absolute Deviation* (MAD) menggunakan rumus sebagai berikut :

$$MAD = \frac{1}{n} \sum_{i=1}^n (|\text{Nilai Forecast} - \text{Nilai Nyata}|)$$

n : jumlah data untuk keperluan peramalan

ZX : jumlah data untuk keperluan peramalan.

2.2.5 Algoritma XGBoost (*eXtreme Gradient Boosting*)

XGBoost adalah sebuah algoritma yang telah ditingkatkan berdasarkan teknik gradient boosting decision tree, yang memungkinkan pembangunan boosted trees dengan efisiensi tinggi dan operasi paralel. Algoritma ini merupakan salah satu metode dalam pembelajaran mesin yang digunakan untuk menangani permasalahan regresi dan klasifikasi dengan menggunakan Gradient Boosting Decision Tree (GBDT). XGBoost pada dasarnya adalah metode ensemble yang berdasarkan pada gradient boosting tree.

Dalam pohon regresi, simpul-simpul internal mewakili nilai-nilai untuk pengujian atribut, sedangkan simpul-simpul daun dengan skor mewakili keputusan. Prediksi akhir diperoleh dengan menjumlahkan skor yang diprediksi oleh setiap pohon K , seperti yang ditunjukkan dalam persamaan berikut:

$$\hat{y} = \sum_k^K f_k(x_i), f_k \in F$$

Metode penelitian mencakup prosedur detail tentang bagaimana penelitian dilakukan. Setiap paragraf dapat terdiri dari beberapa subparagraf yang dijelaskan seperti berikut:

$$obj(\theta) = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_k^K \Omega(f_k)$$

Dimana $\sum_{i=1}^n l(y_i, \hat{y}_i)$ adalah differentiable loss function untuk mengukur apakah model tersebut cocok untuk set data pelatihan dan $\sum_k^K \Omega(f_k)$ adalah item yang menentukan kompleksitas model. Ketika kompleksitas model meningkat skor yang sesuai dikurangi nilainya. Sebelum membangun model prediksi, dilakukan optimasi parameter *turning* XGBoost dengan beberapa parameter [15].

2.2.6 Normalisasi Minmax

Untuk mengurangi kesalahan, dilakukan normalisasi pada kumpulan data dengan mengubah nilai data aktual ke dalam interval range [0,1]. Normalisasi data, juga dikenal sebagai Feature Scaling, adalah proses mengatur struktur data untuk mengurangi sebagian besar ambiguitas. Tujuan normalisasi adalah untuk memastikan bahwa setiap data memiliki bobot yang seimbang [16]. Teknik normalisasi yang diterapkan adalah min-max scaling. Min-Max Normalization adalah metode normalisasi data yang umum digunakan. Ini melibatkan transformasi linier pada data asli untuk menciptakan skala baru [17]. Min-max normalization mempertahankan hubungan antara nilai dalam data

asli dengan rentang yang biasanya ditemui antara 0 hingga 1. Persamaan untuk menghitung normalisasi Min-Max dapat ditemukan dalam bagian persamaan berikut:

[17]

$$x' = \frac{x - \min_x}{\max_x - \min_x}$$

Keterangan:

x' : Hasil normalisasi

x : Data yang akan dinormalisasikan

\min_x : Nilai minimum dari keseluruhan data

\max_x : Nilai maksimum dari keseluruhan data

2.2.7 MAE (Mean Absolute Error)

Mean Absolute Error (MAE) adalah sebuah metode yang digunakan untuk mengevaluasi tingkat ketepatan suatu model peramalan. MAE mengukur rata-rata kesalahan absolut antara nilai peramalan dan nilai yang sebenarnya. Secara lebih spesifik, MAE adalah rata-rata dari nilai absolut dari setiap kesalahan peramalan, di mana kesalahan peramalan merupakan perbedaan absolut antara nilai peramalan dan nilai sebenarnya. MAE adalah ukuran yang umum digunakan untuk menilai tingkat kesalahan prediksi dalam analisis runtun waktu. Semakin rendah nilai MAE, semakin baik model prediksi tersebut dianggap, karena menunjukkan bahwa kesalahan atau error dalam model tersebut cenderung kecil. Oleh karena itu, model prediksi dianggap paling baik ketika nilai MAE mendekati nol. [18].

Secara matematis MAE didefinisikan sebagai berikut:

$$MAE = \frac{1}{n} \sum_{i=1}^n |f_i - y_i|$$

Keterangan:

fi : nilai hasil peramalan

yi : nilai sebenarnya

n : jumlah data.

Berdasarkan formula di atas, MAE secara intuitif menghitung rata – rata error dengan memberikan bobot yang sama untuk seluruh data ($i = 1 \dots n$).

MAE dapat diuraikan menjadi tiga komponen yang mewakili bias (MAE_b), proporsionalitas (MAE_p), dan kesalahan (MAE_M). Keuntungan yang jelas dari dekomposisi kesalahan rata-rata berdasarkan bobot ini adalah bahwa ia menggunakan MAE sebagai garis dasar. Keuntungan lainnya adalah prediksi yang tidak memiliki kesalahan tidak berkontribusi pada komponen. Adapun persamaannya yaitu: [19]

$$MAE_b = \frac{1}{n} \sum_{i=1}^n \frac{b}{b + p_i + u_i} |P_i - O_i|,$$

$$MAE_p = \frac{1}{n} \sum_{i=1}^n \frac{p_i}{b + p_i + u_i} |P_i - O_i|,$$

$$MAE_u = \frac{1}{n} \sum_{i=1}^n \frac{u_i}{b + p_i + u_i} |P_i - O_i|.$$

2.2.8 Root Mean Square Error (RMSE)

Root Mean Square Error (RMSE) adalah suatu metode pengukuran yang digunakan untuk mengevaluasi perbedaan antara nilai prediksi yang dihasilkan oleh sebuah model sebagai perkiraan terhadap nilai yang diamati. RMSE adalah hasil dari akar kuadrat dari Mean Square Error. Tingkat keakuratan metode estimasi kesalahan pengukuran biasanya ditentukan oleh seberapa kecil nilai RMSE yang diperoleh. Metode estimasi yang memiliki nilai Root Mean Square Error (RMSE) yang lebih rendah dianggap lebih akurat daripada metode estimasi yang memiliki nilai RMSE yang lebih tinggi.

Penerapan metode Root Mean Square Error (RMSE) sangat luas, misalnya dalam bidang Meteorologi untuk mengevaluasi model perkiraan matematis tentang lingkungan

atmosfer, dalam bidang ekonomi untuk menilai kecocokan model ekonomi dengan indikator ekonomi, serta dalam ilmu hidrologi untuk mengevaluasi kalibrasi pada model bawah laut. Di industri, RMSE digunakan untuk menilai akurasi metode peramalan, khususnya dalam memprediksi permintaan di masa mendatang.

Langkah-langkah untuk menghitung Root Mean Square Error (RMSE) adalah dengan mengurangkan nilai aktual dengan nilai prediksi, kemudian hasilnya dikuadratkan dan dijumlahkan, lalu hasil penjumlahan tersebut dibagi dengan jumlah data yang ada. Selanjutnya, hasilnya diambil akar kuadratnya untuk mendapatkan nilai RMSE. Rumus untuk menghitung RMSE adalah sebagai berikut:

$$RMSE = \sqrt{\frac{\sum_{t=1}^n (A_t - F_t)^2}{n}}$$

Dimana :

A_t = Nilai data Aktual

F_t = Nilai hasil peramalan

N = banyaknya data

Σ = Summation (Jumlahkan keseluruhan nilai)